A Combinatorial Approach to Nonlocality and Contextuality

Tobias Fritz, Anthony Leverrier, and Ana Belén Sainz

Abstract. So far, most of the literature on (quantum) contextuality and the Kochen-Specker theorem seems either to concern particular examples of contextuality, or be considered as quantum logic. Here, we develop a general formalism for contextuality scenarios based on the combinatorics of hypergraphs which significantly refines a similar recent approach by Cabello, Severini and Winter (CSW). In contrast to CSW, we explicitly include the normalization of probabilities, which gives us a much finer control over the various sets of probabilistic models like classical, quantum and generalized probabilistic. In particular, our framework specializes to (quantum) nonlocality in the case of Bell scenarios, which arise very naturally from the Foulis-Randall product. In the spirit of CSW, we find close relationships to various invariants studied in combinatorics. The recently proposed Local Orthogonality Principle turns out to be a special case of a general principle for contextuality scenarios related to the Shannon capacity of graphs. Our results imply that it is dominated by a low level of the Navascués-Pironio-Acín hierarchy of semidefinite programs, which we apply to contextuality scenarios.

We hope that our approach may also serve as an introduction for combinatorialists to the subject of nonlocality and contextuality. Our conjectures on graphs whose Shannon capacity coincides with their independence number may be of particular interest.

We thank Antonio Acín, Adán Cabello, Simone Severini, Alexander Wilce and Gilles Zémor for comments and discussion, András Salamon for help with a reference, and Will Traves for help on `MathOverflow`. Research at Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Economic Development and Innovation. Part of this work was done while A.L. was at the Institute for Theoretical Physics, ETH Zürich. A.L. was supported by the Swiss National Science Foundation though the National Centre of Competence in Research "Quantum Science and Technology" and through the European Research Council (grant No. 258932). A.B.S. was supported by the ERC SG PERCENT and by the Spanish projects FIS2010-14830, DIQIP and an FPU PhD grant.

Contents

1. Introduction

Much effort has been devoted to understanding the mysteries of quantum theory. In particular, this applies to the phenomena known as quantum **nonlocality** and quantum **contextuality**. Bell's theorem [**Bel64**] shows that no theory can make the same predictions as quantum theory, while jointly satisfying the properties of *realism*, *locality* and *free will*. This is often abbreviated to the statement that quantum theory displays **nonlocality**[1]. Similarly, the Kochen-Specker theorem [**KS67**] states that quantum theory is at variance with any attempt at assigning deterministic values to all observables in a way which would be consistent with the functional relationships between these observables. Such an impossibility is generally known as **contextuality**, since it means that any potential "hidden" value of an observable will necessarily have to depend on the *context* in which it is probed.

It is often stated that nonlocality is, at the mathematical level, a particular case of contextuality. However, it is rarely made explicit what this means precisely. Moreover, the study of contextuality so far often seems to have been concerned with particular examples of contextuality and "small" proofs of the Kochen-Specker theorem, while a general theory has hardly been developed. Some notable exceptions are the following:

(a) The study of **test spaces** in quantum logic [**CMW00**, **Wil09**],
(b) Spekkens' work on **measurement and preparation contextuality** [**Spe05**, **LSW11**],
(c) The **graph-theoretic** approach of Cabello, Severini and Winter [**CSW10**],
(d) The **sheaf-theoretic** approach pioneered by Abramsky and Brandenburger [**AB11**].

Although test spaces are usually considered in the context of quantum logic and state spaces, they serve equally well for the study of contextuality, which is intimately related. This is one of our main themes: a test space can be considered as a **contextuality scenario**, and this is the term we use. As in [**CSW10**], we take a contextuality scenario to be a specification of a collection of measurements which says how many outcomes each measurement has and which measurements have which outcomes in common. We show how the n-party **Bell scenario** with k m-outcome measurements per party can be regarded as a natural contextuality scenarios which one obtains by taking the n-fold **Foulis-Randall product** of the single-party scenario which describes k independent m-outcome measurements. More generally, the Foulis-Randall product is a product operation on contextuality scenarios which naturally incorporates the no-signaling condition. Also, we prove a combinatorial characterization of extremal probabilistic models. In the Bell scenario case, these are the extremal no-signaling boxes.

Our second main theme is to relate, again inspired by CSW, contextuality scenarios and their probabilistic models to graph theory and invariants of graphs like the **independence number**, **Lovász number**, **fractional packing number**. Our approach differs significantly from CSW's in two important respects. First, and most importantly, we explicitly take into account the normalization of probability from the very beginning. In contrast to this, CSW were working with *subnormalized* probabilities, which seems necessary in order to derive their relations to graph-theoretic invariants. We show that such relations still exist even if one retains the normalization of probability. This gives us much finer quantitative information and control about contextuality. Second, while CSW study the maximal values of contextuality **inequalities** for classical, quantum, and general probabilistic models, we consider the sets of classical, quantum, and general probabilistic models themselves as the primary objects. We believe that this is a more natural thing to

[1] This terminology can be confusing, since all known interactions in nature are local [**Haa96**, **Zeh06**], in a different sense of the term.

$$\begin{tikzcd}
\mathcal{C} \arrow[r, phantom, "\overset{5.1.2(b)}{\subseteq}"] \arrow[d, leftrightsquigarrow, "4.3.1"] & \mathcal{Q} \arrow[r, phantom, "\overset{6.2.2}{=}"] & \mathcal{Q}_\infty \arrow[r, phantom, "\subseteq"] & \cdots \arrow[r, phantom, "\subseteq"] & \mathcal{Q}_1 \arrow[r, phantom, "\overset{7.4.1}{\subseteq}"] \arrow[d, leftrightsquigarrow, "6.3.2"] & \mathscr{CE}^\infty \arrow[r, phantom, "\overset{7.3.3}{\subseteq}"] \arrow[d, leftrightsquigarrow, "7.3.2"] & \cdots \arrow[r, phantom, "\overset{7.3.3}{\subseteq}"] & \mathscr{CE}^1 \arrow[r, phantom, "\subseteq"] \arrow[d, leftrightsquigarrow, "7.3.2"] & \mathcal{G} \\
\alpha^* \arrow[rrrr, phantom, "\geqslant"] & & & & \vartheta \arrow[r, phantom, "\geqslant"] & \Theta \arrow[rr, phantom, "\geqslant"] & & \alpha &
\end{tikzcd}$$

FIGURE 1. Chain of inclusions between sets of probabilistic models and corresponding inequalities between graph invariants. We suspect that all inclusions in the first row are strict for some H.

do, since the actual quantities gathered e.g. from an experiment are outcome probabilities rather than coefficients of some inequality. Figure 1 summarizes the sets of probabilistic models that we consider together with their relations to invariants of graphs. The classical $\mathcal{C}$ corresponds to models which can be described in terms of noncontextual hidden variables; $\mathcal{Q}$ is the quantum set; the $\mathcal{Q}_n$ family comes from a hierarchy of semidefinite programs characterizing $\mathcal{Q}$; the general probabilistic set $\mathcal{G}$ contains all models that all conceivable models satisfying the normalization of probability; finally, the family of sets $\mathscr{CE}^n$ arises from our third main theme.

This third main theme is the concept of **Local Orthogonality (LO)** which was recently introduced in [**FSA$^+$12**] as an information-theoretic principle delimiting the set of quantum correlations in Bell scenarios. We show how LO naturally arises in our formalism as a special case of a previously studied concept called **Consistent Exclusivity (CE)** [**Hen12**] or **Specker's Principle** [**Cab12c**]. CE builds on the observation that compatibility of quantum observables is a binary property determined by *pairwise* commutativity. It can be applied both on the single-copy level of a scenario, in which case we write CE^1, and on the many-copy level when the same system is distributed among any number of parties, for which we write CE^∞. This parallels the distinction between LO^1 and LO^∞ that we made in [**FSA$^+$12**]. While CE^1 relates to the **independence number** of a graph, CE^∞ corresponds to the **Shannon capacity** (in the sense of graph theory). This allows us to answer some open questions about LO^∞, though some of our proofs depend on some new conjectures on the Shannon capacity of graphs. In particular, we show that LO^∞, and more generally CE^∞, does not characterize quantum models. In fact, it is satisfied for every probabilistic model which lies in $\mathcal{Q}_1$, which means that a certain semidefinite program is solvable, which (often strictly) contains the quantum set $\mathcal{Q}$. Moreover, at least on some scenarios, there are probabilistic models which satisfy CE^∞, but do not even lie in $\mathcal{Q}_1$. We also ask whether the set of probabilistic models satisfying CE^∞ is convex, and whether activation of CE^∞ violations is possible. These are the questions which reduce to conjectures on the Shannon capacity of graphs. Figure 12 links to all our conjectures and our proofs of implications between them.

1.1. Structure and contents of this paper. We begin in Section 2 by introducing test spaces as our notion of **contextuality scenario**. Later (in Section 3), we will see that every Bell scenario is a contextuality scenario. We continue in Section 2 by defining **probabilistic models** on a contextuality scenario; e.g. for a Bell scenario, these are the no-signaling boxes. We give an abstract characterization of extremal probabilistic models.

In Section 3, we consider products of contextuality scenarios corresponding to simultaneous measurements on spatially separated systems. We find the relevant product operation to be the **Foulis-Randall product** of test spaces. This product guarantees the **no-signaling property** for

probabilistic models on the product scenario by, seemingly paradoxically, incorporating measurements *with* communication. Figure 7 displays the CHSH scenario [**CHSH69**] as a contextuality scenario.

In Section 4, we study **classical models** on contextuality scenarios. These are precisely those probabilistic models that can occur in a world described by noncontextual hidden variables. We introduce the **non-orthogonality graph** of a contextuality scenario and show how its weighted fractional packing number detects the (non-)classicality of a probabilistic model.

In Section 5, we consider **quantum models**. We show how a quantum model of a product of contextuality scenarios arises from quantum representations of the factors such that every operator associated to one factor commutes with an operator associated to another factor.

In Section 6, we show how to formulate a **hierarchy of semidefinite programs characterizing quantum models** for contextuality scenarios. This can be regarded either as a generalization of the original hierarchy for quantum correlations in Bell scenarios [**NPA07**,**NPA08**] or as a special case of the general hierarchy for noncommutative polynomial optimization [**PNA10**]. We relate the first level of this hierarchy to the weighted Lovász number of the non-orthogonality graph.

In Section 7, we consider the principle of Local Orthogonality (LO) introduced in [**FSA$^+$12**] and show in which sense it arises from Consistent Exclusivity (CE) [**Hen12**]. We show how CE relates to the weighted independence number and the weighted Shannon capacity of the non-orthogonality graph. It turns out that the principle, even when applied on the level of distributed copies as CE^∞, is weaker than the first level of the semidefinite hierarchy. We relate the problem of equality and of several other questions about CE^∞, like convexity and activation, to open problems in graph theory. If the non-orthogonality graph is a **perfect graph**, which frequently happens, then every probabilistic model satisfying CE^1 is classical and no interesting contextuality is possible in the given scenario. The strong perfect graph theorem then implies that a scenario can display (quantum) contextuality only if it has a certain odd cycle or odd anti-cycle structure.

In Section 8, we study the complexity of various decision problems on contextuality scenarios. Our "inverse sandwich conjecture" 8.3.3 is an undecidability statement whose proof would have significant repercussions in C^*-algebra theory and quantum logic.

In Section 9, we discuss some further examples of contextuality scenarios and the various sets of probabilistic models associated to them, including a prescription for translating scenarios with subnormalization as in [**CSW10**] into our framework.

In Appendix A, we discuss the graph theory relevant for the main text. In addition to serving as a convenient reference, the main purpose of this is the introduction of our Conjectures A.2.1 on properties of graphs whose Shannon capacity coincides with their independence number.

Finally, in Appendix B, we discuss how our approach, based on hypergraphs in which the vertices represent measurement outcomes, relates to the one of [**AB11**], which is based on hypergraphs in which vertices represent observables. We find that our approach naturally contains the other one.

2. Contextuality scenarios and their probabilistic models

2.1. Motivation: the Kochen-Specker theorem. Cabello *et al.* [**CEGA96**,**Cab08**] showed that one can find 18 vectors in $\mathbb{C}^4$ labeling the vertices of Figure 2 such that the four vectors associated to each one of the 9 edges form an orthonormal basis. Together with the observation that there is no consistent way to label the vertices by $\{0,1\}$ such that every edge contains exactly one vertex labeled by 1, this is a proof of the Kochen-Specker theorem for $\mathbb{C}^4$.

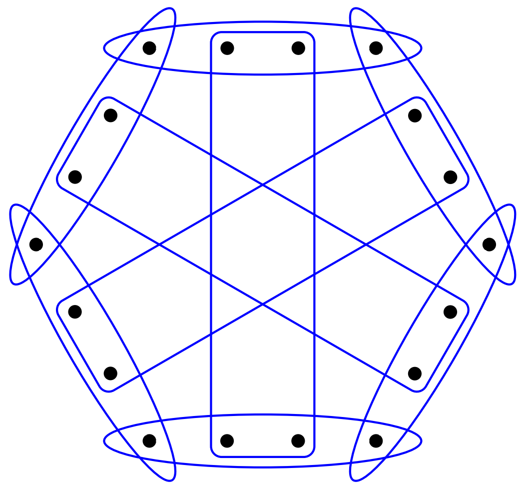


FIGURE 2. The contextuality scenario H_{KS} proving the Kochen-Specker theorem [**CEGA96**, **Cab08**].

Now what does the hypergraph of Figure 2 represent, operationally? This is what we would like to consider next.

2.2. General definition. Since each edge of Figure 2 stands for a basis in $\mathbb{C}^4$, we may think of an edge as representing a 4-outcome measurement. Now every vertex occurs in two different such edges; in other words, the measurements may share outcomes. The assumption of **measurement noncontextuality** [**Spe05**] means that any reasonable theory should represent a shared outcome as a function from states to probabilities which should not depend on the particular measurement in which the outcome occurs.

Abstracting from this particular example to a general definition of **contextuality scenario** means that we need to consider a mathematical structure containing a set of vertices, representing outcomes, and a collection of subsets of the vertices, representing measurements. Mathematically this is a **hypergraph** H with vertices $V(H)$ and edges $E(H)$. We therefore arrive at:

DEFINITION 2.2.1. *A **contextuality scenario** is a hypergraph H such that no edge contains another one:*

$$e_1, e_2 \in E(H),\ e_1 \subseteq e_2 \ \Rightarrow e_1 = e_2, \tag{2.1}$$

and $\bigcup_{e\in E(H)} e = V(H)$.

The reason for postulating (2.1) is related to the normalization of probability: if all outcomes of a measurement e_1 are also outcomes of a measurement e_2, then the additional outcomes of e_2 necessarily have probability 0 and can therefore be disregarded. In the literature on graph theory, hypergraphs satisfying (2.1) are known as **Sperner families** [**Eng97**], or **clutters** [**EF70**]. The condition $\bigcup_{e\in E(H)} e = V(H)$ simply states that each outcome should occur in at least one measurement. In the following, we will generally prefer the term **vertex** over **outcome** and **edge** over **measurement**, while keeping in mind that the latter is the physical interpretation of the former, respectively.

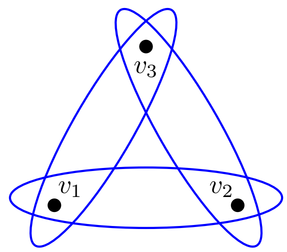


FIGURE 3. The triangle scenario Δ.

In a typical scenario, the hypergraph H is finite, meaning that $V(H)$ is a finite set, and this is the only case that we want to consider.

This kind of definition has been considered before in the literature on contextuality and the Kochen-Specker theorem, e.g. in [**Tka00**, **PMMF10**], and coincides with the notion of **test space** [**Wil09**] which had been introduced in [**FR72**, **RF73**] as **(generalized) sample space**. In particular, the **Greechie diagrams** [**Gre71**, **ST96**] of quantum logic can all be regarded as contextuality scenarios.

On the other hand, Definition 2.2.1 differs from the formalisms proposed in [**AB11**] and [**CF12**, **FC12**]. These works also provide a formalization of contextuality phenomena in terms of hypergraphs, but the vertices of the hypergraph represent observables rather than outcomes, while the edges stand for (maximal) jointly measurable sets of observables. See Appendix B for a more detailed discussion.

2.3. Probabilistic models. In a concrete physical situation, every outcome carries a probability, and the sum of these outcome probabilities over all outcomes in a measurement is 1. This gives:

DEFINITION 2.3.1. *Let H be a contextuality scenario. A* ***probabilistic model*** *on H is an assignment $p : V(H) \to [0,1]$ of a probability $p(v)$ to each vertex $v \in V(H)$ such that*

$$\sum_{v\in e} p(v) = 1 \quad \forall e \in E(H). \tag{2.2}$$

It is important to keep in mind that each $p(v)$ is actually a *conditional* probability: it stands for the probability of getting the outcome v *given that* a measurement $e \ni v$ is conducted.

The set of all probabilistic models on H is a convex subset of $\mathbb{R}^{V(H)}$, possibly empty, which we denote by $\mathcal{G}(H)$. This notation is supposed to suggest the reading "general probabilistic" in the sense of **general probabilistic theories** [**Bar07**]. In the terminology of test spaces [**Wil08**], $\mathcal{G}(H)$ is the set of states over H; unfortunately, the term "probabilistic model" has a different meaning there.

We now turn to some basic examples other than Figure 2. Those mainly interested in Bell scenarios will become satisfied in Section 3.

EXAMPLE 2.3.2. Figure 3 displays the triangle scenario Δ. Its only probabilistic model is $p(v_1) = p(v_2) = p(v_3) = \frac{1}{2}$, since this is the only solution to the system of normalization equations

$$p(v_1) + p(v_2) = 1, \qquad p(v_2) + p(v_3) = 1, \qquad p(v_1) + p(v_3) = 1.$$

See [**LSW11**] for more on this scenario and its unique probabilistic model.

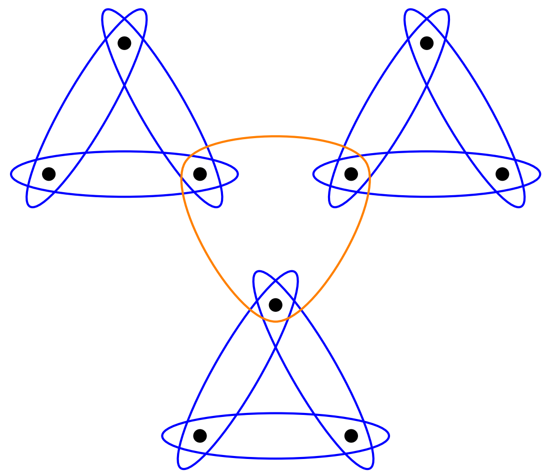


FIGURE 4. Example of a scenario H_0 without any probabilistic model: $\mathcal{G}(H_0) = \varnothing$.

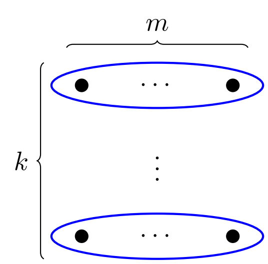


FIGURE 5. The contextuality scenario $B_{1,k,m}$, a "Bell scenario" with only one party.

Contextuality scenarios having a unique probabilistic model, like Δ does, will be of particular importance in Theorem 2.4.3.

EXAMPLE 2.3.3. Figure 4 displays a contextuality scenario H_0 with $\mathcal{G}(H_0) = \varnothing$. Indeed, each of the outer triangles corresponds to a copy of the scenario Δ of Figure 3 and admits a unique probabilistic model where each vertex is assigned a probability 1/2. This is incompatible with the three-outcome measurement depicted in orange which imposes that the probabilities associated with the three corresponding vertices should sum to 1.

EXAMPLE 2.3.4. Figure 5 displays the contextuality scenario defined by k measurements with m outcomes each, such that no two measurements share any outcome. Such scenarios are particularly relevant for describing "box" experiments where an observer can press one of k buttons and record the corresponding measurement outcome.

Further examples will be discussed in Section 9.

For fixed H, the set $\mathcal{G}(H) \subseteq \mathbb{R}^{V(H)}$ is defined in terms of finitely many linear inequalities with rational coefficients. Therefore, it is a convex polytope with rational vertices. A natural question

now is, which polytopes with rational vertices can arise in this way? This has been answered by Shultz:

THEOREM 2.3.5 ([**Shu74**]). *Let $P \subseteq \mathbb{R}^d$ be a polytope with vertices in $\mathbb{Q}^d$. Then there exists a contextuality scenario H_P such that $\mathcal{G}(H_P)$ is affinely isomorphic to P.*

Surprisingly, the combinatorial structure of some polytopes is such that they cannot be represented with rational coordinates only [**Zie95**, Ex. 6.21].

2.4. Characterizing extremal probabilistic models. Since $\mathcal{G}(H)$ is a convex polytope, a natural question is: what are its extreme points? For example, for the CHSH scenario $B_{2,2,2}$ that we will discuss in Section 3, $\mathcal{G}(B_{2,2,2})$ is the no-signaling polytope, and hence its extreme points are the 16 deterministic boxes together with the 8 variants of the PR-box.

In this subsection, we would like to give an abstract characterization of these extremal models which applies to every contextuality scenario.

DEFINITION 2.4.1. *Let H be a contextuality scenario. We say that a non-empty set $W \subseteq V(H)$ **induces a subscenario** if $e_1 \cap W \subseteq e_2 \cap W$ implies that $e_1 = e_2$ for all $e_1, e_2 \in E(H)$. In this case, H_W with*

$$V(H_W) = W; \qquad E(H_W) = \{e \cap W \,:\, e \in E(H)\}$$

*is the **subscenario induced by** W.*

The assumption on W guarantees that H_W is also a contextuality scenario. In particular, it implies that $e \cap W \neq \varnothing$ for all $e \in E(H)$, meaning that W is a **transversal** (or **hitting set**) of the hypergraph H. The subscenario H_W is a **subclutter** of H.

In words: H_W is constructed by dropping all vertices which do not belong to W and restricting all edges accordingly. In doing this, the subset $W \subseteq V(H)$ is assumed to guarantee that no two different edges have equal restrictions or one restriction containing the other.

Intuitively, H_W is the same scenario as H, except that all outcomes not in W have been forbidden. In particular, every probabilistic model p_W on H_W extends to H by setting

$$p(v) \stackrel{def}{=} \begin{cases} p_W(v) & \text{if } v \in W \\ 0 & \text{if } v \notin W \end{cases}.$$

In this case, we say that p is the **extension** of p_W to H.

We have implicitly used induced subscenarios in [**FSA$^+$12**] when considering the graphs on "possible events".

LEMMA 2.4.2. *If H_W is an induced subscenario of H and $H_{W,W'}$ is an induced subscenario of H_W, then $H_{W,W'}$ is also an induced subscenario of H.*

PROOF. Clear. □

Our main result in this section is this:

THEOREM 2.4.3. *$p \in \mathcal{G}(H)$ is extremal if and only if it is the extension of $p_W \in \mathcal{G}(H_W)$ from some induced subscenario H_W which has p_W as its unique probabilistic model.*

PROOF. If H has a unique probabilistic model, i.e. if $\mathcal{G}(H) = \{p\}$, then there is nothing to prove.

Otherwise, the extreme points of $\mathcal{G}(H)$ are precisely the extreme points of the facets of $\mathcal{G}(H)$. Since $\mathcal{G}(H)$ is defined by

$$p(v) \geqslant 0 \ \forall v \in V(H), \qquad \sum_{v \in e} p(v) = 1 \ \forall e \in E(H),$$

for every facet of $\mathcal{G}(H)$ there exists some $v \in V$ such that the facet contains exactly those $p \in \mathcal{G}(H)$ with $p(v) = 0$. We fix such a v and set

$$W = \{w \in V(H) \mid \exists\, p \in \mathcal{G}(H) \text{ s.t. } p(v) = 0 \ \wedge \ p(w) \neq 0\}.$$

In particular, $v \in W$, and W induces a subscenario H_W. By construction, $\mathcal{G}(H_W)$ is the facet of $\mathcal{G}(H)$ defined by $p(v) = 0$.

The assertion then follows by repeatedly applying this process to the induced subscenarios constructed in this way; the lemma guarantees that one obtains an induced subscenario of the original H at each step. This recursion necessarily ends with a scenario which admits a unique probabilistic model, since the dimension of $\mathcal{G}(H)$ decreases by 1 in each step. □

As the proof shows, a similar statement also holds for all faces of $\mathcal{G}(H)$: they all are of the form $\mathcal{G}(H_W)$ for some induced subscenario H_W.

We conclude that an extreme point $p \in \mathcal{G}(H)$ is uniquely determined by the set of vertices $W = \{v \in V(H) \mid p(v) \neq 0\}$, which induces a subscenario H_W with a unique probabilistic model corresponding to forgetting the zeros of p.

The deterministic models of Definition 4.1.1 are a special case of this. Clearly, every deterministic model is an extreme point of $\mathcal{G}(H)$. In terms of Theorem 2.4.3, p is deterministic if and only if each measurement in the associated H_W has only one outcome, i.e. if every vertex in H_W is its own singleton edge. Those extreme points which are not deterministic are the **maximally contextual** models in the scenario H.

3. Products of contextuality scenarios and the no-signaling property

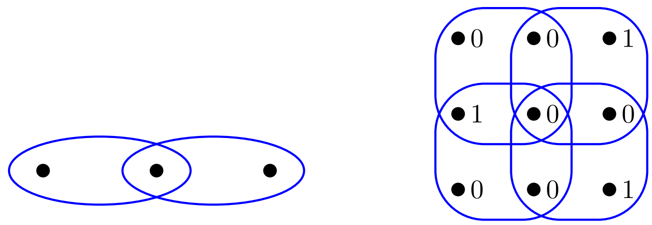
3.1. Products. Imagine two spatially separated or spacelike separated parties Alice and Bob. Alice is assumed to operate in a contextuality scenario H_A, while Bob is taken to operate in a contextuality scenario H_B. Now the two parties can apply simultaneous measurements on their respective systems and will then obtain simultaneous outcomes. In general, the two systems can be correlated, which may lead to correlations between the outcomes. The question now is, what is the contextuality scenario describing this situation? How do two contextuality scenarios combine into a joint one? This question was first answered by Foulis and Randall [FR81], who noticed that the answer is nontrivial. When combining two scenarios into one, we speak of a **product**.

Clearly, the set of outcomes of a product scenario should be the cartesian product of the sets of outcomes, so that a joint outcome simply is the same thing as an outcome of H_A together with an outcome of H_B. Also, every edge on H_A should combine with any edge on H_B into a joint edge. So, naively, one would define the product scenario like this:

Definition 3.1.1. *Let H_A and H_B be contextuality scenarios. The **direct product** is the scenario $H_A \times H_B$ with*

$$V(H_A \times H_B) = V(H_A) \times V(H_B), \qquad E(H_A \times H_B) = E(H_A) \times E(H_B).$$

Now, in any actually observed model, Bob's outcome probabilities should not depend on Alice's choice of measurement and vice versa:



(a) The contextuality scenario Υ.

(b) The direct product $\Upsilon \times \Upsilon$ equipped with a (deterministic) probabilistic model.

FIGURE 6. A contextuality scenario Υ and a probabilistic model on $\Upsilon \times \Upsilon$ not satisfying the no-signaling condition.

DEFINITION 3.1.2. *A probabilistic model $p \in \mathcal{G}(H_A \times H_B)$ is* ***no-signaling*** *if*

(a) For every $w \in V(H_B)$,

$$\sum_{v \in e} p(v, w) = \sum_{v \in e'} p(v, w)$$

for all $e, e' \in E(H_A)$;

(b) For every $v \in V(H_A)$,

$$\sum_{w \in e} p(v, w) = \sum_{w \in e'} p(v, w)$$

for all $e, e' \in E(H_B)$.

This coincides with [**BL09**, Defn. 8] and [**BFRW05**, Defn. 3.2], although the terminology is different.

Now the obvious question is, is every $p \in \mathcal{G}(H_A \times H_B)$ no-signaling? Unfortunately, this is not the case; Figure 6 provides the arguably simplest example. It displays a deterministic model where Alice (vertical) knows with certainty which measurement was performed by Bob (horizontal). It is easy to come up with other examples for virtually any non-trivial scenarios H_A and H_B.

While one solution for this problem is to simply restrict to no-signaling models by fiat [**BL09**], a conceptually much more appealing solution is to use the "right" product of contextuality scenarios:

DEFINITION 3.1.3 ([**FR81**]). *The* ***Foulis-Randall product (FR-product)*** *is the scenario $H_A \otimes H_B$ with*

$$V(H_A \otimes H_B) = V(H_A) \times V(H_B), \qquad E(H_A \otimes H_B) = E_{A \to B} \cup E_{A \leftarrow B}$$

where

$$\begin{aligned} E_{A \to B} &\overset{def}{=} \left\{ \bigcup_{a \in e_A} \{a\} \times f(a) \ : \ e_A \in E_A,\ f : e_A \to E_B \right\}, \\ E_{A \leftarrow B} &\overset{def}{=} \left\{ \bigcup_{b \in e_B} f(b) \times \{b\} \ : \ e_B \in E_B,\ f : e_B \to E_A \right\}. \end{aligned} \tag{3.1}$$

Intuitively, an element of $E_{A \to B}$ is the following: first, an edge $e_A \in E(H_A)$ representing a measurement conducted by Alice; second, a function $f : e_A \to E(H_B)$ which determines the

subsequent measurement of Bob as a function of Alice's outcome. This function f maps each vertex $a \in e_A$ to an edge $f(a) \in E_B$. Similarly for $E_{A \leftarrow B}$, where we think of Bob measuring first and communicating his outcome to Alice, who then chooses her measurement as a function of Bob's outcome. Both possibilities are feasible ways to operate on the joint system and therefore should be considered as measurements conductible on the joint system. In this way, an edge in $H_A \otimes H_B$ is an element of $E_{A\to B}$, $E_{A\leftarrow B}$, or of both sets. For example, Figure 7(f) displays the FR-product of 7(a) with 7(b), which is another copy of 7(a). $E_{A\to B}$ contains the edges of Figure 7(c) and 7(d), while $E_{B\to A}$ consists of 7(c) and 7(e).

Since $H_A \otimes H_B$ contains the same vertices as $H_A \times H_B$ but more edges, we have an inclusion $\mathcal{G}(H_A \otimes H_B) \subseteq \mathcal{G}(H_A \times H_B)$. The following observation is due to Barnum, Fuchs, Renes and Wilce:

Proposition 3.1.4 ([**BFRW05**, Cor. 3.5]). *$\mathcal{G}(H_A \otimes H_B) \subseteq \mathcal{G}(H_A \times H_B)$ is exactly the set of no-signaling models.*

It is in this sense that $H_A \otimes H_B$, in contrast to $H_A \times H_B$, automatically incorporates the no-signaling requirement of special relativity. This is the reason why we regard it as the "right" product of contextuality scenarios.

Both the inclusion $\mathcal{G}(H_A \otimes H_B) \subseteq \mathcal{G}(H_A \times H_B)$ and Proposition 3.1.4 can intuitively be understood in terms of the duality between states and effects [**D'A10**]: restricting the models considered to the no-signaling ones makes more measurements well-defined and in particular allows measurements in which the parties use signaling; on the other hand, allowing measurements in which the parties use signaling is possible only if the system itself, on which the measurements are conducted, does not have internal signaling. Compare Wilce [**Wil08**], who prefers the term *influence-free* over *no-signaling*.

One can also do all this for the case of *unidirectional* no-signaling: defining a product of H_A and H_B by only using the $E_{A\to B}$ of (3.1) gives probabilistic models which are no-signaling from Bob to Alice. See [**BFRW05**] for more details. The resulting product contextuality scenario may be interpreted as describing a temporal succession of operating on H_B after having operated on H_A.

Given two contextuality scenarios H_A and H_B together with probabilistic models

$$p_A \in \mathcal{G}(H_A), \qquad p_B \in \mathcal{G}(H_B),$$

there should exist a probabilistic model $p_A \otimes p_B$ on $H_A \otimes H_B$ having the interpretation of placing physical systems behaving as p_A and p_B "side by side" so that measurements can be conducted on both in parallel, revealing no correlations between the two systems, but independent statistics. To this end, one should obviously define

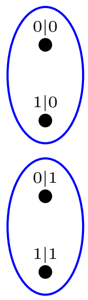
$$p_A \otimes p_B \;:\; V(H_A) \times V(H_B) \longrightarrow [0,1], \qquad (v_A, v_B) \mapsto p_A(v_A) p_B(v_B).$$

Proposition 3.1.5. *This $p_A \otimes p_B$ is a probabilistic model on $H_A \otimes H_B$.*

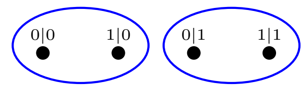
Proof. We need to prove that $\sum_{v\in e} p_A \otimes p_B(v) = 1$ for each edge $e \in E(H_A \otimes H_B)$. Without loss of generality, we can assume $e \in E_{A\to B}$, i.e. $e = \bigcup_{a\in e_A} \{a\} \times f(a)$ for some $e_A \in E_A$ and some $f : e_A \mapsto E_B$, which maps each vertex in e_A to an edge in H_B. Therefore,

$$\sum_{v\in e} p_A \otimes p_B(v) = \sum_{a\in e_A} \sum_{b\in f(a)} p_A(a) p_B(b) = \sum_{a\in e_A} p_A(a) \sum_{b\in f(a)} p_B(b) = \sum_{a\in e_A} p_A(a)\cdot 1 = 1,$$

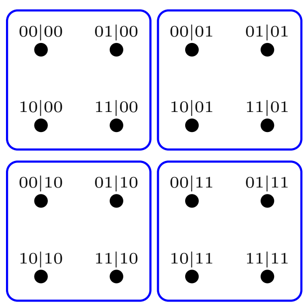
since p_B and p_A are probabilistic models on H_B and H_A, respectively. ☐



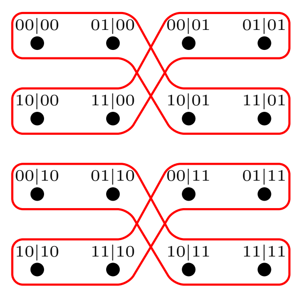
(a) Alice's two binary measurements $B_{1,2,2}$.



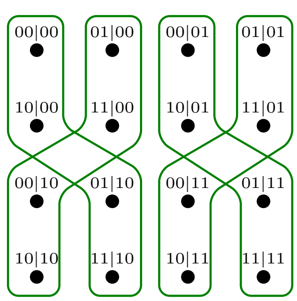
(b) Bob's two binary measurements $B_{1,2,2}$.



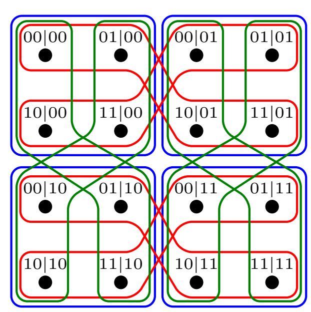
(c) Simultaneous measurements.



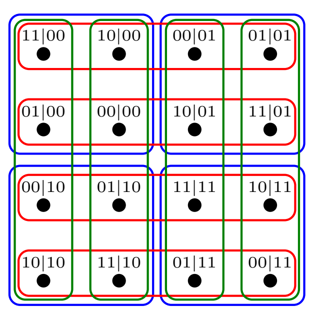
(d) Bob's measurement choice depends on Alice's outcome.



(e) Alice's measurement choice depends on Bob's outcome.



(f) Foulis-Randall product: the CHSH scenario $B_{2,2,2} = B_{1,2,2} \otimes B_{1,2,2}$.



(g) Alternative drawing of 7(f) after rearranging the vertices.

FIGURE 7. Construction of the CHSH scenario $B_{2,2,2}$ as a Foulis-Randall product $B_{2,2,2} = B_{1,2,2} \otimes B_{1,2,2}$.

We write $\mathcal{G}(H_A) \otimes \mathcal{G}(H_B)$ for the set of all probabilistic models of the form $p_A \otimes p_B$. We have just shown that $\mathcal{G}(H_A) \otimes \mathcal{G}(H_B) \subseteq \mathcal{G}(H_A \otimes H_B)$.

REMARK 3.1.6. Often $\mathcal{G}(H_A \otimes H_B)$ is strictly bigger than the convex hull of $\mathcal{G}(H_A) \otimes \mathcal{G}(H_B)$. For example for the Bell scenario $B_{2,2,2} = B_{1,2,2} \otimes B_{1,2,2}$ discussed below, the Popescu-Rohrlich box [**Tsi93**, eq. (1.11)], [**PR94**] is an element of $\mathcal{G}(B_{1,2,2} \otimes B_{1,2,2})$, but does not lie in the convex hull of $\mathcal{G}(B_{1,2,2}) \otimes \mathcal{G}(B_{1,2,2})$.

3.2. Products of more than two scenarios. It is not difficult to check that the Foulis-Randall product "$\otimes$" is a commutative binary operation on contextuality scenarios. But now what about having more than two parties which operate in their respective scenarios simultaneously?

Given three scenarios H_A, H_B, H_C, we can first form the product $H_A \otimes H_B$ and then the product of this with H_C, which gives $(H_A \otimes H_B) \otimes H_C$; unraveling the definitions reveals that an edge in this scenario pertains to one of the four sets

$$E_{(A \to B) \to C}, \qquad E_{(A \leftarrow B) \to C}, \qquad E_{(A \to B) \leftarrow C}, \qquad E_{(A \leftarrow B) \leftarrow C} \tag{3.2}$$

where $E_{(A \to B) \to C}$ is defined to be the collection of all sets of the form

$$\left\{ (a, b, c) \in V(H_A) \times V(H_B) \times V(H_C) \mid a \in e_A,\ b \in f(a),\ c \in g(a, b) \right\} \tag{3.3}$$

where $e_A \in E(H_A)$ is fixed and $f : V(H_A) \to E(H_B)$ and $g : V(H_A) \times V(H_B) \to E(H_C)$ are any functions, and similarly for the other three. In this way, every one of the four sets (3.2) contains all those measurements associated to a certain ordering of the three parties; these four orderings are

$$A \rightsquigarrow B \rightsquigarrow C, \qquad B \rightsquigarrow A \rightsquigarrow C, \qquad C \rightsquigarrow A \rightsquigarrow B, \qquad C \rightsquigarrow B \rightsquigarrow A. \tag{3.4}$$

On the other hand, the bracketing $H_A \otimes (H_B \otimes H_C)$ is based in a similar way on four sets of edges

$$E_{A \to (B \to C)}, \qquad E_{A \leftarrow (B \to C)}, \qquad E_{A \to (B \leftarrow C)}, \qquad E_{A \leftarrow (B \leftarrow C)} \tag{3.5}$$

which represent the time orderings

$$A \rightsquigarrow B \rightsquigarrow C, \qquad B \rightsquigarrow C \rightsquigarrow A, \qquad A \rightsquigarrow C \rightsquigarrow B, \qquad C \rightsquigarrow B \rightsquigarrow A. \tag{3.6}$$

Now these four time orderings are different from (3.4); therefore, in general, $H_A \otimes (H_B \otimes H_C)$ contains different edges than $(H_A \otimes H_B) \otimes H_C$: the Foulis-Randall product is not associative!

Nevertheless, we will frequently want to work with a Foulis-Randall product of more than two factors. One way around the problem of non-associativity is to *define* an n-fold product $H_{A_1} \otimes \ldots \otimes H_{A_n}$ (no brackets) by taking *all* possible time orderings of the parties $A_1, \ldots, A_n$ into account in the sense that an edge in $A_1 \otimes \ldots \otimes A_n$ is defined by the following data:

(a) a permutation σ of the parties such that party $\sigma(t)$ measures at timestep $t \in \{1, \ldots, n\}$;
(b) for every timestep t, a function $f_t : V(H_{\sigma(1)}) \times \ldots \times V(H_{\sigma(t-1)}) \to E(H_{\sigma(t)})$ (possibly constant) which specifies the measurement of party $\sigma(t)$ as a function of all the previous outcomes of parties $\sigma(1), \ldots, \sigma(t-1)$. (For party $\sigma(1)$, this is a function without arguments, i.e. a constant.)

The edge associated to this data is then given by

$$\left\{ (a_1, \ldots, a_n) \mid a_{\sigma^{-1}(i)} \in f_{\sigma^{-1}(i)}(a_{\sigma(1)}, \ldots, a_{\sigma(\sigma^{-1}(i)-1)}) \text{ for } 1 \leqslant i \leqslant n \right\} \subseteq V(H_1) \times \ldots \times V(H_n).$$

For example, in the case of three parties A, B, C, the scenario $H_A \otimes H_B \otimes H_C$ turns out to comprise precisely (3.2) and (3.5). In this way, we obtain an n-ary Foulis-Randall product for any $n \in \mathbb{N}$.

However, for n scenarios $H_{A_1}, \ldots, H_{A_n}$, there are many other ways to form an n-fold product: any consistent way of introducing brackets in the expression

$$H_{A_1} \otimes \ldots \otimes H_{A_n}$$

defines an n-fold Foulis-Randall product. For example,

$$(H_{A_1} \otimes H_{A_2} \otimes H_{A_3}) \otimes H_{A_4} \otimes H_{A_5}$$

represents formation of the ternary product $H_{A_1} \otimes H_{A_2} \otimes H_{A_3}$, followed by taking the ternary product of the resulting scenario with H_{A_4} and H_{A_5}.

Now which one of these products is the "right" one which captures the physical intuition of conducting measurements on independent systems? It turns out that all of these products do the job; while we officially stick to the version without any brackets, all our proofs will also be valid for any non-trivial bracketing; in fact, we suspect that these different products are related via notions like *perspective* [**Wil05**]. In any case, the n-ary product without brackets is "maximal" in the sense that it contains all the edges of all other products. In physical terms, it allows application of the parties' measurements in any temporal order.

3.3. Bell scenarios. We now explain how Bell scenarios are examples of contextuality scenarios. The Bell scenario $B_{n,k,m}$ consists of n parties having access to k measurements each, each of which has m possible outcomes. At the single-party level, the outcomes form a contextuality scenario $B_{1,k,m}$ as depicted in Figure 5. As contextuality scenarios, we define

$$B_{n,k,m} \stackrel{def}{=} \underbrace{B_{1,k,m} \otimes \cdots \otimes B_{1,k,m}}_{n}, \tag{3.7}$$

and we will see in the following how this leads to the usual concepts studied as "nonlocality".

It is straightforward to generalize this definition and all our upcoming results to scenarios where the parties have access to different numbers of measurements and outcomes per measurement, but we will not consider this explicitly.

Example 3.3.1 (The CHSH scenario). Figure 7 illustrate how $B_{2,2,2}$ arises as $B_{1,2,2} \otimes B_{1,2,2}$. A vertex $ab|xy$ represents the event where Alice (resp. Bob) chooses measurement x (resp. y) and obtains output a (resp. b). In this scenario, the edges are as follows:

- For simultaneous measurements, the f of (3.1) are constant, and the measurements are as in Figure 7(c):

$$\{00|00,\ 01|00,\ 10|00,\ 11|00\},$$
$$\{00|01,\ 01|01,\ 10|01,\ 11|01\},$$
$$\{00|10,\ 01|10,\ 10|10,\ 11|10\},$$
$$\{00|11,\ 01|11,\ 10|11,\ 11|11\}.$$

- If Alice measures first and Bob's choice of setting depends on her outcome, then the events are of the form $ab|xf(a)$, where f is not a constant. Thus we have two possibilities: $f(a) = a$ or $f(a) = 1 - a$. In the first case we obtain the edges

$$\{00|00,\ 01|00,\ 10|01,\ 11|01\},$$
$$\{00|10,\ 01|10,\ 10|11,\ 11|11\},$$

and in the second case,

$$\{00|01,\ 01|01,\ 10|00,\ 11|00\},$$
$$\{00|11,\ 01|11,\ 10|10,\ 11|10\}.$$

These are the red edges in Figures 7(d) and 7(f),7(g).

- Similarly, Bob measuring first with Alice's subsequent choice of setting depending on his outcome gives rise to the edges

$$\{00|00,\ 01|10,\ 10|00,\ 11|10\},$$
$$\{00|01,\ 01|11,\ 10|01,\ 11|11\},$$
$$\{00|10,\ 01|00,\ 10|10,\ 11|00\},$$
$$\{00|11,\ 01|01,\ 10|11,\ 11|01\}.$$

These are the green edges in Figures 7(e) and 7(f),7(g).

Proposition 3.3.2. *Let $B_{n,k,m}$ be a Bell scenario. Then $\mathcal{G}(B_{n,k,m})$ is the standard no-signaling polytope containing all no-signaling boxes of type (n,k,m).*

Proof. While this follows from an application of the multipartite version of Proposition 3.1.4, we believe that an independent proof is more instructive.

We identify the vertices of $B_{n,k,m}$ with the events

$$a_1 \dots a_n | x_1 \dots x_n, \quad a_i \in \{1, \dots, m\},\ x_i \in \{1, \dots, k\}$$

in the usual Bell scenario notation.

We show first that a non-signaling box of type (n,k,m) is indeed a probabilistic model on $B_{n,k,m}$. Such a box is an assignment of a probability $p(\vec{a}|\vec{x})$ to each event $\vec{a}|\vec{x}$ such that the no-signaling equations

$$\sum_{a_i} p(a_1 \dots a_n | x_1 \dots x_n) = \sum_{a_i} p(a_1 \dots a_n | x_1 \dots x_i' \dots x_n) \tag{3.8}$$

hold (where the right-hand side is the same except that the setting x_i has been replaced by some other setting x_i'), as well as the normalization condition

$$\sum_{a_1,\dots,a_n} p(a_1 \dots a_n | x_1 \dots x_n) = 1. \tag{3.9}$$

Now we consider any edge in the scenario $B_{n,k,m}$. Without loss of generality, we take the underlying total order of the parties to be the numerical one, so that the temporal order of the parties' measurements is simply $1, \dots, n$. The settings used by the parties are then determined by functions $x_i = f_i(a_1, \dots, a_{i-1})$, and we need to consider

$$\sum_{a_1,\dots,a_n} p(a_1 \dots a_n | f_1() \dots f_n(a_1, \dots a_{n-1})),$$

where $x_1 = f_1()$ is a function without arguments, i.e. a constant. Since the vector of settings does not depend on a_n, the no-signaling equations imply that the last function $f_n(a_1, \dots, a_{n-1})$ can be replaced by an arbitrary constant setting x_n without changing the value of the sum. After applying this modification, the vector of settings does not depend on a_{n-1}, and then the setting of party $n-1$ can be taken to be some fixed x_{n-1}. Applying this procedure repeatedly eventually replaces

all functions $f_i(a_1, \ldots, a_{i-1})$ by constant settings x_i. Then the normalization equation implies that the sum has the value 1, as has been claimed.

Conversely, suppose that p is a probabilistic model on $B_{n,k,m}$. Then p satisfies the normalization equation since taking all functions f_i to be constants x_i gives precisely (3.9). In order to prove the no-signaling equation, we fix arbitrary outputs b_j and choose all functions to be constants $f_j = x_j$, except for

$$f_n(a_1, \ldots, a_{n-1}) = \begin{cases} x_n & \text{if } a_j = b_j \text{ for all } j < n, \\ x_n' & \text{otherwise,} \end{cases}$$

which gives the equation

$$\sum_{a_n} p(b_1 \ldots b_{n-1} a_n | x_1 \ldots x_n) + \sum_{a_n} \sum_{(a_1,\ldots,a_{n-1}) \neq (b_1,\ldots,b_{n-1})} p(a_1 \ldots a_n | x_1 \ldots x_n') = 1.$$

Upon combining this with the already proven normalization equation

$$\sum_{a_n} p(b_1 \ldots b_{n-1} a_n | x_1 \ldots x_n') + \sum_{a_n} \sum_{(a_1,\ldots,a_{n-1}) \neq (b_1,\ldots,b_{n-1})} p(a_1 \ldots a_n | x_1 \ldots x_n') = 1.$$

we obtain (3.8) with $i = n$ and $b_1 \ldots b_{n-1}$ in place of $a_1 \ldots a_{n-1}$. The other no-signaling equations can be obtained in the same way, choosing different orders of the parties. □

In particular, this proof shows explicitly how the non-trivial edges occurring in the definition of "⊗" give rise to the no-signaling property.

4. Classical models

For each scenario H, one can define several relevant subsets of the set of $\mathcal{G}(H)$. In the following, we will define these and study some of their properties in some detail, starting with set of classical models $\mathcal{C}(H)$. We will use the $B_{n,k,m}$ as a "running example" illustrating that $B_{n,k,m}$ indeed behaves exactly as one would expect from the Bell scenario of type (n, k, m).

4.1. Definition. What we mean by **classical** here comprises the idea of hidden variables as they occur in results of Bell [**Bel64**], Fine [**Fin82**] and Kochen-Specker [**KS67**].

Definition 4.1.1. *Let H be a contextuality scenario.*

*(a) A probabilistic model $p \in \mathcal{G}(H)$ is **deterministic** if $p(v) \in \{0, 1\}$ for all $v \in V(H)$.*
*(b) A probabilistic model $p \in \mathcal{G}(H)$ is **classical** if it is a convex combination of deterministic ones.*

Following Fine [**Fin82**] and certain refinements of his results to considerations of contextuality [**LSW11**, Thm. 6], [**AB11**, Thm. 8.1], we note that classical models are precisely those which can be explained in terms of noncontextual hidden variables.

Since, for finite H, there are only finitely many deterministic models, the set of classical models is a polytope. We denote this polytope by $\mathcal{C}(H)$.

Example 4.1.2 ([**CEGA96**]). For H the contextuality scenario of Figure 2, we claim that $\mathcal{C}(H) = \varnothing$. To see this, let V_1 be the set of vertices to which a given deterministic model assigns a 1. Since the set V_1 is required to intersect every edge in precisely one vertex, and every vertex appears in precisely two edges, $2|V_1|$ needs to be equal to the number of edges. Since the latter is odd, we conclude that this is impossible, so that no deterministic model exists, which means that $\mathcal{C}(H) = \varnothing$. See [**AB11**, Sec. 7.1] for a very general version of this argument.

Remark 4.1.3. As we just exemplified, a deterministic model p is determined by the set of vertices

$$V_1 = \{v \in V \mid p(v) = 1\}\,. \tag{4.1}$$

By definition of deterministic model, V_1 has the property that it intersects every edge in exactly one vertex: V_1 is an **exact transversal** [**Eit94**]. Conversely, every exact transversal V_1 defines a deterministic model in this way. We have that $\mathcal{C}(H) \neq \varnothing$ if and only if H has an exact transversal.

In the same way that probalistic models coincide with the non-signaling polytopes, the classicality of a contextuality scenario naturally extends that of Bell scenarios.

Proposition 4.1.4. *Let $B_{n,k,m}$ be a Bell scenario. Then $\mathcal{C}(B_{n,k,m})$ is the standard Bell polytope.*

Proof. This is clear since one way to define the Bell polytope is as the convex hull of deterministic models, and a deterministic model in the contextuality scenario $B_{n,k,m}$ is the same as a local deterministic model in the Bell sense. □

4.2. Non-orthogonality graphs. We will now start to relate contextuality scenarios and the associated classes of probabilistic models to graph theory.

In the (hyper-)graph theory literature, one frequently considers the **orthogonality graph** of a hypergraph (also referred to as its **primal** or **Gaifman graph** [**GLS01**]). The orthogonality graph of a hypergraph H is obtained by replacing all edges in H by complete subgraphs on the same vertices. This coincides with the orthogonality relation present in the **generalized sample spaces** of [**FR72**]. Moreover, upon thinking of H as an abstract simplicial complex with E as its facets, its orthogonality graph is the 1-skeleton of this simplicial complex.

For the purpose of relating to the graph theory of Appendix A, however, it will be more convenient to consider the complement of this graph. The drawback of this is that it makes some of our considerations a bit more confusing, like the proof of Lemma 4.2.2.

Definition 4.2.1 (Non-orthogonality graph). *Let H be a contextuality scenario. The **non-orthogonality graph** $\mathrm{Ort}(H)$ is the undirected graph with the same vertices as H and adjacency relation*

$$u \sim v \iff \nexists e \in E(H) \text{ with } \{u, v\} \subseteq e.$$

We say that two different vertices u and v of H are orthogonal, which we denote by $u \perp v$, if they are not adjacent in $\mathrm{Ort}(H)$, i.e. if they do belong to a common edge in H. We now make use of the concepts discussed in Appendix A, starting with the strong product of graphs $\boxtimes$.

Lemma 4.2.2. *Let H_A and H_B be contextuality scenarios. Then,*

$$\mathrm{Ort}(H_A \otimes H_B) = \mathrm{Ort}(H_A) \boxtimes \mathrm{Ort}(H_B).$$

Proof. Clearly both sides are graphs having $V(H_A) \times V(H_B)$ as their set of vertices, so what needs to be shown is that the adjacency relations coincide.

We first prove that if $(u_A, u_B) \perp (v_A, v_B)$ in $\mathrm{Ort}(H_A \otimes H_B)$, then these two vertices are also not adjacent in $\mathrm{Ort}(H_A) \boxtimes \mathrm{Ort}(H_B)$. The assumption means that there is an edge $e \in E(H_A \otimes H_B)$ which contains both (u_A, u_B) and (v_A, v_B); this edge has one of the two forms of (3.1). If it is in $E_{A\to B}$, then $u_A, v_A \in e_A$, meaning that $u_A \perp v_A$. Similarly, if the edge is in $E_{A\leftarrow B}$, then $u_B \perp v_B$. The conclusion follows from either case.

For proving the opposite implication, we show that $(u_A, u_B) \perp (v_A, v_B)$ in $\mathrm{Ort}(H_A) \boxtimes \mathrm{Ort}(H_B)$ implies the same in $\mathrm{Ort}(H_A \otimes H_B)$. The assumption means that $u_A \perp v_A$ or $u_B \perp v_B$; by symmetry,

it is enough to consider the case $u_A \perp v_A$. Then, there exists some $e_A \in E(H_A)$ with $u_A, v_A \in e_A$. Now choose $e_B, e'_B \in E_B$ such that $u_B \in e_B$ and $v_B \in e'_B$, and some function $f : e_A \to E_B$ with $f(u_A) = e_B$ and $f(v_A) = e'_B$. Then

$$\bigcup_{a \in e_A} \{a\} \times f(a)$$

is an edge in $H_A \otimes H_B$ containing (u_A, u_B) and (v_A, v_B), which proves the claim. □

4.3. Classicality from the fractional packing number. We now show how to detect classicality using a graph-theoretic invariant from Appendix A.

PROPOSITION 4.3.1. *A probabilistic model $p \in \mathcal{G}(H)$ is in $\mathcal{C}(H)$ if and only if $\alpha^*(\mathrm{Ort}(H), p) \leqslant 1$.*

Note that the normalization $\sum_{v \in e} p(v) = 1$ for every $e \in E(H)$ implies that $\alpha^*(\mathrm{Ort}(H), p) \geqslant 1$, so that the condition $\alpha^*(\mathrm{Ort}(H), p) \leqslant 1$ is actually equivalent to $\alpha^*(\mathrm{Ort}(H), p) = 1$.

PROOF. By definition, $\alpha^*(\mathrm{Ort}(H), p) \leqslant 1$ means that if $q : V(H) \to [0, 1]$ are vertex weights satisfying $\sum_{v \in C} q_v \leqslant 1$ for all cliques $C \subseteq \mathrm{Ort}(H)$, then also

$$\sum_{v \in V(H)} q_v p(v) \leqslant 1. \tag{4.2}$$

In order to prove the claim for all classical p, it is sufficient to consider deterministic p. In this case, the associated set $V_1 = \{v \in V(H) \mid p(v) = 1\}$ is itself a clique in $\mathrm{Ort}(H)$, while all other $p(v)$ vanish, and hence (4.2) follows from the assumption on q.

For the other direction, we use the dual formulation (A.7) of the weighted fractional packing number: there exists a number $x_C \geqslant 0$ associated to every clique $C \subseteq \mathrm{Ort}(H)$ such that $p(v) \leqslant \sum_{C \ni v} x_C$ and $\sum_C x_C = 1$. We claim that every C for which $x_C \neq 0$ corresponds to a deterministic model via (4.1); in other words, if $x_C \neq 0$, then $|e \cap C| = 1$ for every $e \in E(H)$. First, $|e \cap C| \leqslant 1$, since e is an independent set in $\mathrm{Ort}(H)$ while C is a clique. Second, the chain of inequalities

$$1 = \sum_{v \in e} p(v) \leqslant \sum_{v \in e} \sum_{C \ni v} x_C = \sum_{C \text{ with } C \cap e \neq \varnothing} x_C \leqslant \sum_C x_C = 1$$

actually needs to be a chain of equalities, which proves the claim that if $x_C \neq 0$, then $|e \cap C| = 1$ for every $e \in E(H)$. Furthermore, we also conclude that $p(v) = \sum_{C \ni v} x_C$, or $p = \sum_C x_C \mathbb{1}_C$. This is an explicit decomposition of p as a convex combination of deterministic models. □

PROBLEM 4.3.2. Can this result be used to derive a combinatorial characterization of the facets of $\mathcal{C}(H)$, similar in spirit to Theorem 2.4.3?

4.4. Classical models on products.

PROPOSITION 4.4.1.

$$\mathcal{C}(H_A \otimes H_B) = \mathrm{conv}\left(\mathcal{C}(H_A) \otimes \mathcal{C}(H_B)\right),$$

where conv(S) *denotes the convex hull of the elements in* S.

This is supposed to be seen in contrast to Remark 3.1.6.

PROOF. Let $p_A \in \mathcal{C}(H_A)$ and $p_B \in \mathcal{C}(H_B)$ be deterministic models. Then also $p_A \otimes p_B$ is a deterministic model on $H_A \otimes H_B$, which proves $\mathcal{C}(H_A \otimes H_B) \supseteq \mathrm{conv}\left(\mathcal{C}(H_A) \otimes \mathcal{C}(H_B)\right)$ by convexity of $\mathcal{C}(H_A \otimes H_B)$.

Conversely, consider a deterministic model p_{AB} on $H_A \otimes H_B$. Let V_1 be the set of vertices in $H_A \otimes H_B$ for which $p_{AB}(v) = 1$, and define $p_A \in \mathcal{C}(H_A)$ and $p_B \in \mathcal{C}(H_B)$ as follows: for each

$v_A \in V_A$, set $p_A(v_A) = 1$ if and only if there exists $v_B \in V_B$ such that $(v_A, v_B) \in V_1$, and $p_A(v_A) = 0$ otherwise. Similarly, define p_B. We want to check that these are indeed probabilistic models, i.e. show that $\sum_{v_A \in e_A} p_A(v_A) = 1$ and $\sum_{v_B \in e_B} p_B(v_B) = 1$ for every edge e_A of H_A and e_B of H_B. As V_1 is an exact transversal of $H_A \otimes H_B$, no two elements of V_1 belong to the same edge. This implies that if both $(u_A, u_B), (u'_A, u'_B) \in V_1$, then there is no $e_A \in E(H_A)$ with $\{u_A, u'_A\} \subseteq e_A$: for if there where, then we could construct an edge as in the proof of Lemma 4.2.2 which contains both (u_A, u_B) and (u'_A, u'_B). It follows for each edge $e_A \in E_A$, there is at most one vertex $v_A \in e_A$ with $p_A(v_A) = 1$. In fact, there is exactly one such vertex, since $e_A \times e_B$ is an edge on $H_A \otimes H_B$ for any $e_B \in E(H_B)$, and this edge intersects with V_1. Hence, p_A is a probabilistic model on H_A. Similarly, p_B is a probabilistic model. Since $p_{AB} = p_A \otimes p_B$ by construction, the claim follows by convexity. □

5. Quantum models

5.1. Definition and basic properties. We denote by $\mathcal{B}(\mathcal{H})$ the set of all bounded operators on a Hilbert space $\mathcal{H}$. The notation $\mathcal{B}_+(\mathcal{H})$ stands for the subset of positive semi-definite operators. A quantum state ρ is given by a normalized density operator, i.e. by some $\rho \in \mathcal{B}_{+,1}(\mathcal{H})$, where $\mathcal{B}_{+,1}(\mathcal{H}) \stackrel{def}{=} \{\rho \in \mathcal{B}_+(\mathcal{H}) \mid \operatorname{tr} \rho = 1\}$.

Definition 5.1.1. *Let H be a contextuality scenario. An assignment of probabilities $p : V(H) \to [0,1]$ is a **quantum model** if there exists a Hilbert space $\mathcal{H}$, a quantum state $\rho \in \mathcal{B}_{+,1}(\mathcal{H})$ and a projection operator $P_v \in \mathcal{B}(\mathcal{H})$ associated to every $v \in V$ which constitute projective measurements in the sense that*

$$\sum_{v \in e} P_v = \mathbb{1}_{\mathcal{H}} \quad \forall e \in E(H), \tag{5.1}$$

and reproduce the given probabilities,

$$p(v) = \operatorname{tr}(\rho P_v) \quad \forall v \in V(H). \tag{5.2}$$

The set of all quantum models is the **quantum set** $\mathcal{Q}(H)$. Thanks to (5.1), it is clear that $\mathcal{Q}(H) \subseteq \mathcal{G}(H)$, i.e. every quantum model is a probabilistic model.

Proposition 5.1.2. *(a) $\mathcal{Q}(H)$ is convex.*
(b) Every classical model is a quantum model: $\mathcal{C}(H) \subseteq \mathcal{Q}(H)$.

Proof. (a) Let $p_1, p_2 \in \mathcal{Q}(H)$ be quantum models implemented by Hilbert spaces $\mathcal{H}_1$, $\mathcal{H}_2$, projection operators $P_{1,v}$, $P_{2,v}$ and states ρ_1, ρ_2 on the respective Hilbert space. Then for any coefficient $\lambda \in [0,1]$, we construct a quantum representation of $\lambda p_1 + (1-\lambda)p_2$ by setting

$$\mathcal{H} \stackrel{def}{=} \mathcal{H}_1 \oplus \mathcal{H}_2, \qquad P_v \stackrel{def}{=} P_{1,v} \oplus P_{2,v}, \qquad \rho \stackrel{def}{=} \lambda\rho_1 \oplus (1-\lambda)\rho_2.$$

It is immediate to verify that this is indeed a quantum representation of $\lambda p_1 + (1-\lambda)p_2$.

(b) This follows from (a) upon showing that every deterministic model is quantum. A deterministic model p can be seen to be quantum by setting $\mathcal{H} = \mathbb{C}$, $P_v = p(v) \cdot \mathbb{1}$ and $\rho = \mathbb{1}$.

□

5.2. Quantum models on products. What is the set of quantum models on a product $H_A \otimes H_B$? The following characterization generalizes the **commutativity paradigm** of quantum correlations in Bell scenarios [**JNP$^+$11**, **Fri12**]. For a related argument, see [**CSW10**, (iv)].

PROPOSITION 5.2.1. *Let H_A and H_B be two contextuality scenarios. Then $p \in \mathcal{Q}(H_A \otimes H_B)$ if and only if it there is a Hilbert space $\mathcal{H}$, a quantum state $\rho \in \mathcal{B}_{+,1}(\mathcal{H})$ and projection operators $P_{A,u} \in \mathcal{B}(\mathcal{H})$, $P_{B,v} \in \mathcal{B}(\mathcal{H})$ assigned to every $u \in V(H_A)$, $v \in V(H_B)$ such that*

$$\sum_{u \in e_A} P_{A,u} = \mathbb{1}_{\mathcal{H}} = \sum_{v \in e_B} P_{B,v} \quad \forall e_A \in E(H_A),\ e_B \in E(H_B),$$

$$[P_{A,u}, P_{B,v}] = 0 \qquad \forall u \in V(H_A),\ v \in V(H_B),$$

and the given probabilistic model is reproduced,

$$p(u,v) = \mathrm{tr}\,(\rho\, P_{A,u} P_{B,v}) \qquad \forall u \in V(H_A),\ v \in V(H_B). \tag{5.3}$$

PROOF. We start from (5.3) and assign to every vertex $(u,v) \in V(H_A \otimes H_B)$ the projection

$$P_{(u,v)} \stackrel{def}{=} P_{A,u} P_{B,v},$$

so that (5.2) holds by (5.3). By symmetry, it is sufficient to show (5.1) for an edge $e \in E_{A \to B}$ given by

$$e \stackrel{def}{=} \bigcup_{a \in e_A} \{a\} \times f(a) \text{ with } e_A \in E_A,\ f : e_A \to E_B.$$

In this case,

$$\sum_{w \in e} P_w = \sum_{u \in e_A} P_{A,u} \sum_{v \in f(u)} P_{B,v} = \sum_{u \in e_A} P_{A,u} \cdot \mathbb{1}_{\mathcal{H}} = \mathbb{1}_{\mathcal{H}},$$

which is analogous to the computation in the proof of Proposition 3.1.5.

Conversely, one can construct the "local" observables $P_{A,u}$ and $P_{B,v}$ from a quantum model on $\mathcal{Q}(H_A \otimes H_B)$ by noting that the operators

$$P_{A,u} \stackrel{def}{=} \sum_{v \in e_B} P_{(u,v)}, \qquad P_v \stackrel{def}{=} \sum_{u \in e_A} P_{(u,v)} \tag{5.4}$$

do not depend on the choice of $e_B \in E(H_B)$ or $e_A \in E(H_A)$, respectively. To see this, it is enough to prove that

$$\sum_{v \in e_B} P_{(u,v)} = \sum_{v \in e'_B} P_{(u,v)} \tag{5.5}$$

for any $u \in V(H_A)$ and $e_B, e'_B \in E(H_B)$, which is analogous to the proof of Proposition 3.3.2. Choosing some $e_A \ni u$ and considering the function $f : e_A \to E(H_B)$ with

$$f(u') = \begin{cases} e_B & \text{if } u' = u, \\ e'_B & \text{otherwise.} \end{cases}$$

An application of (5.1) to the edge defined by f as well as the edge $e_A \times e'_B$ gives

$$\sum_{v \in e_B} P_{(u,v)} + \sum_{u' \in e_A \setminus \{u\}} \sum_{v \in e'_B} P_{(u',v)} = \mathbb{1}_{\mathcal{H}} = \sum_{u' \in e_A} \sum_{v \in e'_B} P_{(u',v)},$$

which reduces to (5.5) after cancelling terms. This shows that the "local" operators (5.4) are well-defined.

The normalization condition $\sum_u P_{A,u} = \mathbb{1}_{\mathcal{H}} = \sum_v P_{B,v}$ now is an immediate consequence of (5.1). Finally, the commutativity $[P_{A,u}, P_{B,v}] = 0$ follows again from the normalization

$$\sum_{u' \in e_A} \sum_{v' \in e_B} P_{(u',v')} = \mathbb{1}_{\mathcal{H}},$$

taken for some $e_A \ni u$ and $e_B \ni v$: the terms in this sum are necessarily mutually orthogonal, and hence commute pairwise; but now both $P_{A,u}$ and $P_{B,v}$ are partial sums of this big sum, and therefore these commute as well. Also, mutual orthogonality implies $P_{(u,v)} = P_{A,u} P_{B,v}$, which yields the desired probabilities (5.3). □

For example, $\mathcal{H}$ might itself be a tensor product $\mathcal{H}_A \otimes \mathcal{H}_B$, such that every $P_{A,u}$ operates on the first factor, while every $P_{B,v}$ operates on the second, while ρ is a state on $\mathcal{H}_A \otimes \mathcal{H}_B$, possibly entangled. The question whether every quantum model on $\mathcal{Q}(H_A \otimes H_B)$ arises, at least approximately, from this **tensor paradigm** is a generalization of Tsirelson's problem [**JNP+11**, **Fri12**].

Problem 5.2.2. Is the set of all quantum models in the tensor paradigm dense in $\mathcal{Q}(H_A \otimes H_B)$?

There are some immediate consequences of Proposition 5.2.1:

Corollary 5.2.3.

$$\mathcal{Q}(H_A) \otimes \mathcal{Q}(H_B) \subseteq \mathcal{Q}(H_A \otimes H_B) \tag{5.6}$$

Again, the CHSH scenario $B_{2,2,2} = B_{1,2,2} \otimes B_{1,2,2}$ exemplifies that (5.6) is not an equality in general. A proof analogous to Proposition 5.2.1 also holds for quantum models on n-fold products, and from this we deduce that:

Corollary 5.2.4. *$\mathcal{Q}(B_{n,k,m})$ is the set of quantum correlations in the Bell sense in the commutativity paradigm.*

5.3. Quantum contextuality. We conclude this section by formulating the Kochen-Specker theorem of [**CEGA96**] in our formalism: the classical set of the scenario H_{KS} of Figure 2 is empty, while the quantum set is nonempty:

Theorem 5.3.1 (Kochen-Specker). *There exists a contextuality scenario H_{KS} for which*

$$\mathcal{C}(H_{KS}) = \varnothing, \qquad \mathcal{Q}(H_{KS}) \neq \varnothing.$$

It is not clear to us whether $\mathcal{Q}(H_{KS}) = \mathcal{G}(H_{KS})$, but we suspect that this is not the case. A natural question is whether there exists a proof of the Kochen-Specker in which this is the case:

Problem 5.3.2. Is there H for which

$$\mathcal{C}(H) \neq \mathcal{Q}(H) = \mathcal{G}(H) \quad ?$$

Some hypergraph H constructed from the GHZ paradox [**GHSZ90**] might be a good candidate for this hypothetical phenomenon.

Proposition 5.3.3. *There exists H as in Problem 5.3.2 if and only there exists some H' with a unique probabilistic model which is quantum, but not classical.*

Proof. Clearly if such an H' exists, then we can take $H = H'$ in Problem 5.3.2. Conversely, such an H' can be constructed as an induced subscenario of any H of Problem 5.3.2 by using Theorem 2.4.3, whose proof adapts immediately to show that the resulting unique probabilistic model H' will also be quantum. □

6. A hierarchy of semidefinite programs characterizing quantum models

For Bell scenarios, there is a sequence ("hierarchy") of semidefinite programs characterizing quantum correlations with the commutativity paradigm due to Navascués, Pironio and Acín [**NPA07**, **NPA08**]. Here, we extend this hierarchy to contextuality scenarios. This may be considered a special case of the general hierarchy for noncommutative polynomial optimization [**PNA10**].

6.1. Definition of the hierarchy. We introduce the main idea before getting to the technical details. Given a quantum model as in Definition 5.1.1, not only can one consider the expectation values $\mathrm{tr}\,(\rho P_v)$, but also any expectation value of the form

$$\mathrm{tr}\,(\rho P_{v_1} \dots P_{v_n})\,, \tag{6.1}$$

where $\mathbf{v} = v_1 \dots v_n \in V(H)^n$ is any finite sequence of vertices. The idea is to find properties of these values which characterize quantum models. For $n \geqslant 2$, these values are typically not determined by the probabilities $p(v) = \mathrm{tr}\,(\rho P_v)$ alone; the hierarchy works with these quantities as unknown variables whose values have to be determined in such a way as to be consistent with arising from a quantum model as in (6.1).

Now for some notation. As a shorthand, we also write $P_{\mathbf{v}}$ for $P_{v_1} \dots P_{v_n}$, although this is in general not a projection. When V is a set, we write V^{*n} for the set of all strings of up to n elements of V, i.e. $V^{*n} = \bigcup_{k\leqslant n} V^k$, and $V^* = \bigcup_{k\in\mathbb{N}} V^k$ for the set of all strings over V. $\varnothing \in V^*$ is the empty string of length 0 and the associated operator is $P_\varnothing \stackrel{def}{=} \mathbb{1}$. For $\mathbf{v} = v_1 \dots v_n$ a string, we denote its reverse by $\mathbf{v}^\dagger = v_n \dots v_1$. This notation makes sense in our context since $P_{\mathbf{v}^\dagger} = P_{\mathbf{v}}^\dagger$. For $\mathbf{v} \in V^*$ and $\mathbf{w} \in V^*$, we write their concatenation simply as $\mathbf{vw} \in V^*$, so that $P_{\mathbf{vw}} = P_{\mathbf{v}} P_{\mathbf{w}}$. We also use $v_1 \dots \hat{v}_i \dots v_n$ as a shorthand for $v_1 \dots v_{i-1} v_{i+1} \dots v_n$.

LEMMA 6.1.1. *Let $p \in \mathcal{Q}(H)$ be a quantum model with $v \mapsto P_v \in \mathcal{B}_+(\mathcal{H})$, $\rho \in \mathcal{B}_{+,1}(\mathcal{H})$. Then the matrix M indexed by $\mathbf{v}, \mathbf{w} \in V(H)^{*n}$ with entries*

$$M_{\mathbf{v},\mathbf{w}} = \mathrm{tr}\left(\rho P_{\mathbf{v}} P_{\mathbf{w}}^\dagger\right) \tag{6.2}$$

has the following properties:

(a) M is positive semidefinite.

(b)

$$M_{\varnothing,\varnothing} = 1. \tag{6.3}$$

(c) For every $e \in E(H)$,

$$\sum_{x\in e} M_{\mathbf{v}x,\mathbf{w}} = M_{\mathbf{v},\mathbf{w}}. \tag{6.4}$$

(d) If $v_n \perp w_m$, then

$$M_{v_1\dots v_n, w_1\dots w_m} = 0. \tag{6.5}$$

Hermiticity of M implies that (6.4) also holds with x appended to $\mathbf{w}$ rather than $\mathbf{v}$.

PROOF. (a) It needs to be shown that for any vector $x \in \mathbb{C}^{V(H)^{*n}}$ with components $x_{\mathbf{v}} \in \mathbb{C}$, $\mathbf{v} \in V(H)^{*n}$, the expression

$$\sum_{\mathbf{v},\mathbf{w}} x_{\mathbf{v}}^* M_{\mathbf{v},\mathbf{w}} x_{\mathbf{w}}.$$

is nonnegative. By the definition (6.2), this is equal to

$$\sum_{\mathbf{v},\mathbf{w}} \mathrm{tr}\left(\rho\, x_{\mathbf{v}}^* P_{\mathbf{v}} P_{\mathbf{w}}^\dagger x_{\mathbf{w}}\right).$$

With $Q = \sum_{\mathbf{v}} x_{\mathbf{v}} P_{\mathbf{v}}^\dagger$, this is of the form $\operatorname{tr}\left(\rho Q^\dagger Q\right)$, and therefore indeed nonnegative.

(b) Since ρ is a normalized state, $M_{\varnothing,\varnothing} = \operatorname{tr}(\rho) = 1$.

(c) The requirement (5.1) implies that

$$\sum_{x \in e} P_{\mathbf{v}x} = P_{\mathbf{v}},$$

from which (6.4) directly follows.

(d) This is a direct consequence of $P_{v_n} \perp P_{w_m}$ for $v_n \perp w_m$.

□

The diagonal entries $M_{\mathbf{v},\mathbf{v}}$ represent the expectation values $\operatorname{tr}\left(\rho P_{v_1} \dots P_{v_n} P_{v_n} \dots P_{v_1}\right)$ which can be interpreted as the probability to obtain the sequence of outcomes $(v_1, \dots, v_n)$ in some measurement sequence $(e_1, \dots, e_n)$ with $v_i \in e_i\ \forall i$. We suspect that this interpretation can be used to find an interpretation of the "higher" levels of the hierarchy in terms of the lowest level of a temporally extended scenario, but we have not been able to make this idea work.

Definition 6.1.2. *Let H be a contextuality scenario. We say that $p : V(H) \to [0,1]$ is a $\mathcal{Q}_n$-model if there is a positive semidefinite matrix M with entries $M_{\mathbf{v},\mathbf{w}}$, $\mathbf{v},\mathbf{w} \in V(H)^{*n}$, such that (6.3), (6.4), (6.5) hold and*

$$p(v) = M_{v,\varnothing}. \tag{6.6}$$

Again, it is easy to verify that every $\mathcal{Q}_n$-model is a probabilistic model. By definition, testing whether a given probabilistic model lies in $\mathcal{Q}_n$ is a semidefinite programming problem of size $|V(H)|^n \times |V(H)|^n$. By making judicious use of the equations (6.4) and the upcoming (6.10), this size can be significantly reduced if H has many edges; any practical computation should take this into account. Furthermore, it can be assumed that all matrix entries are actually in $\mathbb{R}$, i.e. no imaginary components are needed.

Remark 6.1.3. Besides those of Lemma 6.1.1, there are other properties satisfied by M which follow from (6.3)–(6.5), and are satisfied in particular by those M of the form (6.2):

(a) If $\mathbf{v}\mathbf{w}^\dagger = \mathbf{v}'\mathbf{w}'^\dagger$, then

$$M_{\mathbf{v},\mathbf{w}} = M_{\mathbf{v}',\mathbf{w}'}. \tag{6.7}$$

This follows by induction from $M_{v_1 \dots v_m,\mathbf{w}} = M_{v_1 \dots v_{m-1},\mathbf{w}v_m}$, which in turn is a consequence of (6.4) and (6.5) upon choosing some $e \ni v_m$,

$$M_{v_1 \dots v_m,\mathbf{w}} \overset{(6.4)}{=} \sum_{x \in e} M_{v_1 \dots v_m,\mathbf{w}x} \overset{(6.5)}{=} M_{v_1 \dots v_m,\mathbf{w}v_m},$$

and applying the same trick on the other side shows that this also equals $M_{v_1 \dots v_{m-1},\mathbf{w}}$, as claimed. Equation (6.7) implies in particular that all matrix entries $M_{\mathbf{v},\mathbf{w}}$ are determined by those of the "first row", i.e. those of the form $M_{\varnothing,\mathbf{v}}$, although this requires $\mathbf{v} \in V(H)^{*2n}$.

(b) Repeating one letter in the index string gives the same matrix entry,

$$M_{v_1 \dots v_i \dots v_m,\mathbf{w}} = M_{v_1 \dots v_i v_i \dots v_m,\mathbf{w}}. \tag{6.8}$$

Upon using (6.7), this follows from a very similar argument.

(c) For every $e \in E(H)$,

$$\sum_{v_i \in e} M_{\mathbf{v},\mathbf{w}} = M_{v_1 \dots \hat{v}_i \dots v_m,\mathbf{w}}. \tag{6.9}$$

This is a consequence of (6.4) and (6.7).

(d) Having subsequent orthogonal indices makes the matrix entry vanish,

$$v_i \perp v_{i+1} \quad \Longrightarrow \quad M_{v_1 \ldots v_i v_{i+1} \ldots v_m, \mathbf{w}} = 0. \tag{6.10}$$

This follows from (6.9) together with (6.8).

(e) Choosing some $e \ni v$ and applying (6.4) and (6.5) also shows that

$$M_{v,\varnothing} = M_{v,v}. \tag{6.11}$$

In particular, $p(v) = M_{v,v}$ by (6.6).

Definition 6.1.2 is our semidefinite hierarchy for contextuality scenarios. In the special case of a bipartite Bell scenario $B_{2,k,m}$, our hierarchy is equivalent to the original one [**NPA07**, **NPA08**], although our level "n" is somewhat different from the hierarchy level "n" used in [**NPA08**]. In particular, our set $\mathcal{Q}_1(B_{2,k,m})$ is the set Q^{1+AB} of [**NPA08**].

PROPOSITION 6.1.4.

$$\mathcal{Q}(H) \subseteq \ldots \subseteq \mathcal{Q}_n(H) \subseteq \ldots \subseteq \mathcal{Q}_1(H).$$

PROOF. Every matrix M showing that p is a $\mathcal{Q}_{n+1}$-model can be restricted to a matrix showing that p is a $\mathcal{Q}_n$-model, so that $\mathcal{Q}_{n+1}(H) \subseteq \mathcal{Q}_n(H)$. Lemma 6.1.1 shows that every quantum model is a $\mathcal{Q}_n$-model, which means that $\mathcal{Q}(H) \subseteq \mathcal{Q}_n(H)$. □

6.2. Convergence of the hierarchy. We can also consider infinite matrices M with entries $M_{\mathbf{v},\mathbf{w}}$ indexed by strings of arbitrary length $\mathbf{v}, \mathbf{w} \in V(H)^*$; starting from a quantum model and considering (6.2) as the resulting definition of the matrix, the same proof as before shows that the properties of Lemma 6.1.1 still hold, if we take positive semidefiniteness to mean that

$$\sum_{\mathbf{v},\mathbf{w} \in V(H)^*} x_{\mathbf{v}}^* M_{\mathbf{v},\mathbf{w}} x_{\mathbf{w}} \geqslant 0$$

for all finitely supported $(x_{\mathbf{v}})_{\mathbf{v} \in V(H)^*}$.

PROPOSITION 6.2.1. *If such an infinite matrix exists, then* $p \in \mathcal{Q}$.

PROOF. Such an infinite matrix M can be understood to be a ($*$-algebraic) state ϕ on the $*$-algebra with generators $\{P_v,\ v \in V(H)\}$ and relations

$$P_v = P_v^2 = P_v^*, \qquad \sum_{v \in e} P_v = \mathbb{1} \quad \forall e \in E(H) \tag{6.12}$$

via the assignment

$$\phi\left(P_{v_1} \ldots P_{v_n}\right) \stackrel{def}{=} M_{v_1 \ldots v_n, \varnothing}.$$

and extending by linearity. Then, the GNS construction (see e.g. [**KR83**]) turns this into a quantum representation satisfying (6.6). For this reason, a probabilistic model is quantum if and only if there exists such an infinite matrix M having the properties of Lemma 6.1.1.

More concretely, this works as follows. First, we claim that

$$\sum_{\mathbf{v},\mathbf{w} \in V(H)^*} x_{\mathbf{v}}^* M_{\mathbf{v}u,\mathbf{w}u} x_{\mathbf{w}} \leqslant \sum_{\mathbf{v},\mathbf{w} \in V(H)^*} x_{\mathbf{v}}^* M_{\mathbf{v},\mathbf{w}} x_{\mathbf{w}}. \tag{6.13}$$

To see this, choose any $e \ni u$ and write

$$\begin{aligned}\sum_{\mathbf{v},\mathbf{w}\in V(H)^*} x^*_{\mathbf{v}} \left(M_{\mathbf{v},\mathbf{w}} - M_{\mathbf{v}u,\mathbf{w}u}\right) x_{\mathbf{w}} &\overset{6.1.3}{=} \sum_{\mathbf{v},\mathbf{w}\in V(H)^*} x^*_{\mathbf{v}} \left(\sum_{u'\in e,\, u'\neq u} M_{\mathbf{v}u',\mathbf{w}u'} \right) x_{\mathbf{w}} \\ &= \sum_{u'\in e,\, u'\neq u} \; \sum_{\mathbf{v},\mathbf{w}\in V(H)^*} x^*_{\mathbf{v}} M_{\mathbf{v}u',\mathbf{w}u'} x_{\mathbf{w}} \geqslant 0,\end{aligned}$$

where the last inequality is due to positive semidefiniteness of M. This proves (6.13).

Now we start the construction with the infinite-dimensional vector space spanned by all strings, $\mathcal{H}_0 \overset{def}{=} \mathrm{lin}_{\mathbb{C}}\left(V(H)^*\right)$. The formula

$$\left\langle \sum_{\mathbf{v}\in V(H)^*} x_{\mathbf{v}}\mathbf{v}, \sum_{\mathbf{w}\in V(H)^*} x_{\mathbf{w}}\mathbf{w} \right\rangle \overset{def}{=} \sum_{\mathbf{u},\mathbf{v}\in V(H)^*} x^*_{\mathbf{u}} M_{\mathbf{v},\mathbf{w}} x_{\mathbf{v}}.$$

defines a positive semidefinite inner product on $\mathcal{H}_0$ in terms of the matrix M. The Cauchy-Schwarz inequality shows that

$$\mathcal{N} \overset{def}{=} \left\{ \sum_{\mathbf{v}\in V(H)^*} x_{\mathbf{v}}\mathbf{v} \in \mathcal{H}_0 \;\middle|\; \left\langle \sum_{\mathbf{v}} x_{\mathbf{v}}\mathbf{v}, \sum_{\mathbf{v}} x_{\mathbf{v}}\mathbf{v} \right\rangle = 0 \right\}$$

is a linear subspace of $\mathcal{H}_0$. The inner product on the quotient $\mathcal{H}_0/\mathcal{N}$ then is positive definite by definition. We take $\mathcal{H}$ to be the completion of $\mathcal{H}_0/\mathcal{N}$ with respect to the norm coming from this inner product.

Now for $u \in V(H)$, the operator P_u is defined to act on $\mathcal{H}_0$ as

$$P_u \left(\sum_{\mathbf{v}\in V(H)^*} x_{\mathbf{v}}\mathbf{v} \right) \overset{def}{=} \sum_{\mathbf{v}\in V(H)^*} x_{\mathbf{v}}\mathbf{v}u.$$

Thanks to (6.13), this descends to a well-defined operator in $\mathcal{B}(\mathcal{H})$, which we also denote by P_u. The equation $M_{\mathbf{v}u,\mathbf{w}} = M_{\mathbf{v},\mathbf{w}u}$ guarantees that P_u is self-adjoint, while $M_{\mathbf{v}uu,\mathbf{w}} = M_{\mathbf{v}u,\mathbf{w}}$ shows that $P_u^2 = P_u$ since

$$\sum_{\mathbf{v}\in V(H)^*} x_{\mathbf{v}} \left(\mathbf{v}uu - \mathbf{v}u\right) \in \mathcal{N}.$$

The equation $\sum_{u\in e} P_u = \mathbb{1}_{\mathcal{H}}$ holds since

$$\sum_{\mathbf{v}\in V(H)^*} x_{\mathbf{v}} \left(\mathbf{v} - \sum_{u\in e} \mathbf{v}u \right) \in \mathcal{N}$$

thanks to (6.4). Finally, the rank-one density operator associated to the empty string $\varnothing \in \mathcal{H}$ is the desired quantum state, since

$$\langle \varnothing, P_u \varnothing \rangle = M_{\varnothing,u} = p(u).$$

This ends the GNS construction. □

From this reasoning, we find that the sequence of sets $(\mathcal{Q}_n)_{n\in\mathbb{N}}$ converges in the following sense:

Theorem 6.2.2.

$$\mathcal{Q} = \bigcap_{n\in\mathbb{N}} \mathcal{Q}_n.$$

Proof ([**NPA08**]). It needs to be shown that if $p \in \mathcal{Q}_n$ for all $n \in \mathbb{N}$, then $p \in \mathcal{Q}$. To this end, we show that if a matrix $(M^n_{\mathbf{v},\mathbf{w}})_{\mathbf{v},\mathbf{w}\in V(H)^{*n}}$ exists with the required properties for every n, then there also exists a corresponding infinite matrix $(M^\infty_{\mathbf{v},\mathbf{w}})_{\mathbf{v},\mathbf{w}\in V(H)^*}$.

For $\mathbf{v} \in V(H)^{*n}$, positive semidefiniteness gives the estimate

$$\left(M^{2n}_{\mathbf{v},\mathbf{v}}\right)^2 \overset{(6.7)}{=} \left(M^{2n}_{\mathbf{v}\mathbf{v}^\dagger,\varnothing}\right)^2 \leqslant M^{2n}_{\mathbf{v}\mathbf{v}^\dagger,\varnothing} \cdot M^{2n}_{\varnothing,\varnothing} = M^{2n}_{\mathbf{v},\mathbf{v}},$$

which implies $M^{2n}_{\mathbf{v},\mathbf{v}} \leqslant 1$, and hence

$$|M^{2n}_{\mathbf{v},\mathbf{w}}|^2 \leqslant M^{2n}_{\mathbf{v},\mathbf{v}} M^{2n}_{\mathbf{w},\mathbf{w}} \leqslant 1$$

again thanks to positive semidefiniteness. We obtain $M^k_{\mathbf{v},\mathbf{w}} \in [-1,+1]$ for all $\mathbf{v},\mathbf{w} \in V(H)^{*n}$ with $n \leqslant 2k$.

Now consider the truncation of any M^{2n} to a matrix indexed by $\mathbf{v},\mathbf{w} \in V(H)^{*n}$. Upon filling this truncation up with 0's, we obtain an infinite matrix indexed by $\mathbf{v},\mathbf{w} \in V(H)^*$ with all elements in $[-1,+1]$. In this way, every matrix M^n becomes an element of $[-1,+1]^{V(H)^*\times V(H)^*}$. The space $[-1,+1]^{V(H)^*\times V(H)^*}$, equipped with the product topology, is second countable, and also compact thanks to Tychonoff's theorem. Hence, the sequence $(M^n)_{n\in\mathbb{N}}$ has a convergent subsequence, and we write M^∞ for its limit. By construction, this M^∞ is an infinite matrix indexed by $\mathbf{v},\mathbf{w} \in V(H)^*$ having all the desired properties. The claim now follows from Proposition 6.2.1. ☐

Since each $\mathcal{Q}_n(H)$ is defined in terms of a semidefinite program, we say that this represents a **hierarchy of semidefinite programs** characterizing $\mathcal{Q}(H)$. It is a subfamily of the hierarchies of semidefinite programs in noncommutative optimization introduced in [**PNA10**], which generalize the "commutative" hierarchies originally discovered in the context of convex optimization [**Las02**].

6.3. Equivalent definitions of $\mathcal{Q}_1$ and the Lovász number. The following is a reformulation of the set "$\mathcal{E}^1_{\mathrm{QM}}$" considered in [**CSW10**].

Proposition 6.3.1. *For $p \in \mathcal{G}(H)$, the following are equivalent:*

(a) $p \in \mathcal{Q}_1(H)$;

(b) There exists a Hilbert space $\mathcal{H}$, a unit vector $|\Psi\rangle \in \mathcal{H}$ and a vector $|\phi_v\rangle$ for every $v \in V(H)$ such that
 - *(i)* $u \perp v \implies \langle\phi_u|\phi_v\rangle = 0$,
 - *(ii)* $\sum_{v\in e} |\phi_v\rangle = |\Psi\rangle \quad \forall e \in E(H)$,
 - *(iii)* $p(v) = \langle\phi_v|\phi_v\rangle$;

(c) There exists a Hilbert space $\mathcal{H}$, a unit vector $|\Psi\rangle \in \mathcal{H}$ and a unit vector $|\psi_v\rangle$ for every $v \in V(H)$ such that
 - *(i)* $u \perp v \implies \langle\psi_u|\psi_v\rangle = 0$,
 - *(ii)* $p(v) = |\langle\psi_v|\Psi\rangle|^2$;

(d) There exists a Hilbert space $\mathcal{H}$, a unit vector $|\Psi\rangle \in \mathcal{H}$ and a projection P_v for every $v \in V(H)$ such that
 - *(i)* $u \perp v \implies P_u \perp P_v$,
 - *(ii)* $p(v) = \langle\Psi|P_v|\Psi\rangle \quad \forall v \in V(H)$;

(e) There exists a Hilbert space $\mathcal{H}$, a unit vector $|\Psi\rangle \in \mathcal{H}$ and a projection P_v for every $v \in V(H)$ such that
 - *(i)* $\sum_{v\in e} P_v \leqslant \mathbb{1}_{\mathcal{H}} \quad \forall e \in E(H)$,
 - *(ii)* $p(v) = \langle\Psi|P_v|\Psi\rangle \quad \forall v \in V(H)$;

In all cases, $\mathcal{H}$ can also be taken to be the real Hilbert space $\mathbb{R}^{|V(H)|}$.

PROOF. (a)⇒(b): By positive semidefiniteness, we can write M as a Gram matrix, so that there exist vectors $|\Psi\rangle, |\phi_v\rangle$ in $\mathcal{H} = \mathbb{R}^{|V(H)|}$ such that

$$M_{\varnothing,\varnothing} = \langle\Psi|\Psi\rangle, \qquad M_{\varnothing,v} = \langle\Psi|\phi_v\rangle, \qquad M_{u,v} = \langle\phi_u|\phi_v\rangle,$$

from which (b)(i) and (b)(iii) follow.

Now we fix $e \in E(H)$ and show (b)(ii). We decompose $|\Psi\rangle$ into orthogonal components $|\Psi\rangle = |\Psi^{\|}\rangle + |\Psi^{\perp}\rangle$, where $|\Psi^{\|}\rangle \in \mathrm{lin}_{\mathbb{C}}\{|\phi_v\rangle \,:\, v \in e\}$. Due to (6.4) and (6.5), the vectors satisfy:

$$\langle\phi_v|\Psi\rangle = M_{v,\varnothing} = \sum_{u\in e} M_{v,u} = M_{v,v} = \langle\phi_v|\phi_v\rangle.$$

Then the equations

$$\langle\phi_v|\Psi\rangle = \langle\phi_v|\phi_v\rangle$$

imply that $|\Psi^{\|}\rangle = \sum_{v\in e}|\phi_v\rangle$. On the other hand,

$$\langle\Psi^{\|}|\Psi^{\|}\rangle + \langle\Psi^{\perp}|\Psi^{\perp}\rangle = M_{\varnothing,\varnothing} = \sum_{v\in e} M_{\varnothing,v} = \sum_{v,u\in e} M_{u,v} = \sum_{v,u\in e}\langle\phi_u|\phi_v\rangle = \langle\Psi^{\|}|\Psi^{\|}\rangle$$

shows that $|\Psi^{\perp}\rangle = 0$, so that $\sum_{v\in e}|\phi_v\rangle = |\Psi\rangle$, as desired.

(b)⇒(c): Normalizing the $|\phi_v\rangle$ to $|\psi_v\rangle \stackrel{def}{=} \frac{1}{\sqrt{\langle\phi_v|\phi_v\rangle}}|\phi_v\rangle$ guarantees the orthogonality relations, and choosing some edge $e \in E(H)$ with $v \in e$ gives

$$|\langle\psi_v|\Psi\rangle|^2 = \frac{1}{\langle\phi_v|\phi_v\rangle}\left|\left\langle \phi_v \middle| \sum_{u\in e}\phi_u \right\rangle\right|^2 = \frac{1}{\langle\phi_v|\phi_v\rangle}\langle\phi_v|\phi_v\rangle^2 = \langle\phi_v|\phi_v\rangle,$$

due to the orthogonality relations.

(c)⇒(d): Define $P_v = |\psi_v\rangle\langle\psi_v|$.

(d)⇒(e): This is clear since for fixed $e \in E(H)$, all projections P_v for $v \in e$ are mutually orthogonal, which implies $\sum_{v\in e} P_v \leqslant \mathbb{1}_{\mathcal{H}}$.

(e)⇒(a): Define $M_{v,w} = \langle\Psi|P_v P_w^{\dagger}|\Psi\rangle$. We check that M satisfies conditions (6.3) to (6.5) and is positive semidefinite:

(6.3) $M_{\varnothing,\varnothing} = \langle\Psi|\Psi\rangle = 1$, since $|\Psi\rangle$ is a unit vector.

(6.4) Consider an edge $e \in E$. Since $p(v)$ is a probabilistic model,

$$\langle\Psi|\Psi\rangle = 1 = \sum_{v\in e} p(v) = \langle\Psi|\sum_{v\in e} P_v|\Psi\rangle,$$

which implies $\sum_{v\in e} P_v|\Psi\rangle = |\Psi\rangle$. Then,

$$\sum_{v\in e} M_{v,w} = \langle\Psi|\sum_{v\in e} P_v P_w|\Psi\rangle = \langle\Psi|P_w|\Psi\rangle = M_{\varnothing,w}.$$

(6.5) If $v \perp w$, then there is an edge $e \in E(H)$ with $v, w \in e$. Hence, $P_v \perp P_w$, so that $M_{v,w} = \langle\Psi|P_v P_w|\Psi\rangle = 0$.

Positive semidefiniteness of M can be shown as in the proof of Lemma 6.1.1.

□

PROPOSITION 6.3.2. *A probabilistic model $p \in \mathcal{G}(H)$ is in $\mathcal{Q}_1$ if and only if $\vartheta(\mathrm{Ort}(H), p) \leqslant 1$.*

Proof. We use the characterization of $\mathcal{Q}_1(H)$ given in Proposition 6.3.1(c). Assuming $p \in \mathcal{Q}_1(H)$, we choose corresponding vectors $|\psi_v\rangle, |\Psi\rangle \in \mathbb{R}^{|V(H)|}$; then, by Definition A.3.1,

$$\vartheta(\mathrm{Ort}(H), p) \leqslant \max_{v\in V} \frac{p(v)}{|\langle\Psi|\psi_v\rangle|^2} = \frac{p(v)}{p(v)} = 1.$$

Conversely, if $\vartheta(\mathrm{Ort}(H), p) \leqslant 1$, then there is an orthonormal labeling $(|\psi_v\rangle)_{v\in V(H)}$ and a vector $|\Psi\rangle \in \mathbb{R}^{|V|}$ such that $|\langle\Psi|\psi_v\rangle|^2 \geqslant p(v)\ \forall v$. By choosing $\mathcal{H} = \mathbb{R}^{|V(H)|} \oplus \mathbb{R}^{|V(H)|}$ and setting

$$|\psi'_v\rangle \stackrel{def}{=} \frac{\sqrt{p(v)}}{|\langle\Psi|\psi_v\rangle|}|\psi_v\rangle \oplus \sqrt{1 - \frac{p(v)}{|\langle\Psi|\psi_v\rangle|^2}}|e_v\rangle \in \mathcal{H}$$

where the $|e_v\rangle$ form the standard basis of $\mathbb{R}^{|V(H)|}$, one obtains $|\langle\Psi|\psi'_v\rangle|^2 = p(v)$ with unit vectors $|\psi'_v\rangle$, as desired. □

This relation to graph theory has a simple first application:

Proposition 6.3.3.

$$\mathcal{Q}_1(H_A) \otimes \mathcal{Q}_1(H_B) \subseteq \mathcal{Q}_1(H_A \otimes H_B) \tag{6.14}$$

Proof. Combine Proposition 6.3.2 with multiplicativity of ϑ (Proposition A.3.10). □

Again, the CHSH scenario $B_{2,2,2} = B_{1,2,2} \otimes B_{1,2,2}$ exemplifies that (6.14) is not an equality in general, even after taking the convex hull on the left-hand side.

7. Consistent Exclusivity and Local Orthogonality

7.1. Introducing Consistent Exclusivity. It is a fundamental property of quantum theory that the compatibility of observables is a *binary* relation: if a collection of quantum observables is such that they commute pairwise, then it follows that there is a basis in which all of them are diagonal, so that a measurement in that basis can be coarse-grained into a measurement of each observable. Paraphrasing Specker [**Spe60**][2],

> A collection of propositions about a quantum mechanical system is precisely then simultaneously decidable, when they are pairwise simultaneously decidable.

For us, this means the following: suppose that $I \subseteq V(H)$ is a set of vertices in a contextuality scenario H such that every two of them belong to a common edge; by definition of $\mathrm{Ort}(H)$, this means precisely that I is an independent set in $\mathrm{Ort}(H)$. Then the associated projections $(P_v)_{v\in I}$ for any quantum model $p \in \mathcal{Q}(H)$ have the property that of being pairwise orthogonal, and hence $\sum_{v\in I} P_v \leqslant \mathbb{1}_{\mathcal{H}}$. This implies

$$\sum_{v\in I} p(v) = \sum_{v\in I} \mathrm{tr}(\rho P_v) \leqslant 1.$$

Definition 7.1.1 ([**Hen12**]). *A probabilistic model $p \in \mathcal{G}(H)$ satisfies **Consistent Exclusivity** if*

$$\sum_{v\in I} p(v) \leqslant 1 \tag{7.1}$$

holds for any independent set $I \subseteq V(\mathrm{Ort}(H))$. We write $\mathcal{CE}^1(H) \subseteq \mathcal{G}(H)$ for the set of probabilistic models satisfying CE.

[2]We thank Adán Cabello for pointing us to [**Spe60**] and kindly sharing a preliminary version of [**Cab12c**].

We also write CE^1 for this version of CE in order to distinguish it from the upcoming refinement termed CE^∞. We refer to [**Cab12c**] for an exposition of the history of principle and in which contexts it has been applied.

Intuitively, CE is saying that the total probability of any collection of pairwise exclusive outcomes is $\leqslant 1$. In this formulation, CE may almost sound like a trivial consequence of the laws of probability; however, this is not the case, since the probabilities $p(v)$ of a probabilistic model are *conditional* probabilities representing the probability that outcome v occurs *given that* a measurement e with $v \in e$ has been performed.

Proposition 7.1.2. *(a) $\mathcal{Q}(H) \subseteq \mathscr{CE}^1(H)$ for every H.*
(b) There exist H with $\mathscr{CE}^1(H) \subsetneq \mathcal{G}(H)$.

Proof. (a) Above.
(b) For the triangle scenario Δ of Figure 3, $V(\Delta)$ is itself an independent set in $\mathrm{Ort}(\Delta)$. Since $\sum_{v\in V(\Delta)} p(v) = \frac{3}{2}$ for the unique probabilistic model p, this model violates CE. We conclude that $\mathscr{CE}^1(\Delta) = \varnothing$, although $\mathcal{G}(\Delta) = \{p\}$.

See [**LSW11**] for further discussion of this example and [**FSA$^+$12**] for examples in multipartite Bell scenarios.

□

In [**CSW10**], Consistent Exclusivity was imposed in the very *definition* of probabilistic models. The problem with this is that the collection of models satisfying it is not closed under $\otimes$, as we will see in the following. Aside from the unclear physical meaning of CE, this is the main reason why we prefer our Definition 2.3.1: it guarantees that if p_A and p_B are probabilistic models on H_A and H_B, respectively, then $p_A \otimes p_B$ is also a probabilistic model on $H_A \otimes H_B$; see Section 3.1.

By the very Definition A.3.1 of the weighted independence number, we have:

Proposition 7.1.3. *$p \in \mathscr{CE}^1(H)$ if and only if $\alpha(\mathrm{Ort}(H), p) \leqslant 1$.*

Again, due to the normalization equations $\sum_{v\in e} p(v) = 1$, the statement $\alpha(\mathrm{Ort}(H), p) \leqslant 1$ is actually equivalent to $\alpha(\mathrm{Ort}(H), p) = 1$.

7.2. Local Orthogonality. The concept of **Local Orthogonality (LO)** was recently introduced in [**FSA$^+$12**] as an information-theoretic principle satisfied by all quantum correlations in Bell scenarios, but violated by many non-quantum no-signaling boxes. The main reason for considering LO is the search for "physical" principles characterizing quantum correlations. It seems intuitively related to Consistent Exclusivity; here we would like to explain in which sense it is indeed a special case of CE when using our definition (3.7) of Bell scenario.

Recall [**FSA$^+$12**] that we call two events $u = a_1 \dots a_n | x_1 \dots x_n$ and $v = a'_1 \dots a'_n | x'_1 \dots x'_n$ in a Bell scenario **locally orthogonal** if there is a party i with $a_i \neq a'_i$, but $x_i = x'_i$. We now show that two events are locally orthogonal if and only if they are different vertices belonging to a common edge in the hypergraph $B_{n,k,m}$:

Lemma 7.2.1. *The events $u, v \in V(B_{n,k,m})$ are locally orthogonal if and only if $u \perp v$.*

Proof. Suppose that $u = a_1 \dots a_n | x_1 \dots x_n$ and $v = a'_1 \dots a'_n | x'_1 \dots x'_n$ are locally orthogonal. By relabeling the parties, we can arrange for $a_1 \neq a'_1$ and $x_1 = x'_1$. Now choose any functions $f_2, \dots, f_n$ with $f_i(a_1) = x_i$ and $f_i(a'_1) = x'_i$. Then the set of events of the form

$$b_1 \dots b_n | x_1 f_2(b_1) \dots f_n(b_1)$$

defines an edge in $B_{n,k,m}$ containing both u and v. Intuitively, Alice communicates her outcome to the other parties who then choose their measurement settings as a function of that outcome.

Conversely, $u \perp v$ means that there is an edge $e \in E(B_{n,k,m})$ with $u, v \in e$. More concretely, this states that there is an ordering of the parties $\sigma(1), \ldots, \sigma(n)$ and functions $f_{\sigma(i)}(b_{\sigma(1)}, \ldots, b_{\sigma(i-1)})$ such that e contains exactly those events which have the form

$$b_{\sigma(1)} \ldots b_{\sigma(n)} | f_{\sigma(1)}() \ldots f_{\sigma(n)}(b_{\sigma(1)}, \ldots, b_{\sigma(n-1)})$$

where we have now written the parties in the order given by the permutation σ. Since both given events $u = a_1 \ldots a_n | x_1 \ldots x_n$ and $v = a'_1 \ldots a'_n | x'_1 \ldots x'_n$ are assumed to be of this form, we know that $x_{\sigma(i)} = f_{\sigma(i)}(a_{\sigma(1)}, \ldots, a_{\sigma(i-1)})$ and $x'_{\sigma(i)} = f_{\sigma(i)}(a'_{\sigma(1)}, \ldots, a'_{\sigma(i-1)})$. Now let $\sigma(j)$ be the smallest index with $a_{\sigma(j)} \neq a'_{\sigma(j)}$. Then, since $x_{\sigma(j)}$ and $x'_{\sigma(j)}$ only depend on $a_{\sigma(i)}$ and $a'_{\sigma(i)}$ with $i < j$, we conclude that $x_{\sigma(j)} = x'_{\sigma(j)}$, which proves the claim. □

Hence, when working within our framework for contextuality scenarios, the LO^1 principle studied in [**FSA$^+$12**] becomes a special case of CE^1 of Definition 7.1.1; the orthogonality between two events naturally arises from the FR product. Those readers not familiar with [**FSA$^+$12**] may regard this as the definition of LO^1. In [**Cab12b**], this relation between LO^1 and CE^1 was already implicitly used.

Problem 7.2.2. In [**FSA$^+$12**], we have introduced LO^1, which we now know to define $\mathscr{CE}^1(B_{n,k,m})$, as a principle limiting information processing power. Can this characterization of $\mathscr{CE}^1$ be extended to all contextuality scenarios?

Problem 7.2.3. In [**FSA$^+$12**], we also showed that LO^1 is equivalent to the no-signaling principle in bipartite Bell scenarios, i.e. $\mathscr{CE}^1(B_{2,k,m}) = \mathcal{G}(B_{2,k,m})$. More generally, under which conditions on H does $\mathscr{CE}^1(H) = \mathcal{G}(H)$ hold?

7.3. Consistent Exclusivity and the Shannon capacity of graphs. If $p \in \mathcal{G}(H)$ is a probabilistic model which is realizable in a world obeying certain physical laws, then it is reasonable to assume that any $p^{\otimes n} \in \mathcal{G}(H^{\otimes n})$ is realizable as well, since it simply corresponds to conducting n copies of the same experiment in parallel. If we regard CE as delimiting the set of physically realizable probabilistic models, then this means that if $p^{\otimes n} \notin \mathscr{CE}^1(H^{\otimes n})$, then we already know that p itself is not physically realizable. Therefore we put:

Definition 7.3.1 (CE hierarchy of sets). *Let H be a contextuality scenario and $p \in \mathcal{G}(H)$. We write $p \in \mathscr{CE}^n(H)$ if and only if $p^{\otimes n} \in \mathscr{CE}^1(H^{\otimes n})$. Furthermore,*

$$\mathscr{CE}^\infty(H) \stackrel{def}{=} \bigcap_{n \in \mathbb{N}} \mathscr{CE}^n(H).$$

This is indeed relevant since, as we saw in [**FSA$^+$12**], for example $\mathscr{CE}^2(B_{2,2,2}) \neq \mathscr{CE}^1(B_{2,2,2})$. See [**Cab12b**] for another example showing that violations of CE can be "activated" by considering copies $p^{\otimes n}$ of the same model p. If $p \in \mathscr{CE}^k(H)$, then we also say that p satisfies CE^k. In particular, $p \in \mathscr{CE}^\infty(H)$ if and only if $p \in \mathscr{CE}^n(H)$ for all $n \in \mathbb{N}$, in which case we say that p satisfies CE$^\infty$. In the special case of Bell scenarios, our previous results imply that CE$^\infty$ is precisely LO$^\infty$ of [**FSA$^+$12**].

We now relate the $\mathscr{CE}^*$ family of sets to the graph-theoretical invariants of Appendix A.

Lemma 7.3.2. *(a) $p \in \mathscr{CE}^n(H)$ if and only if $\alpha(\mathrm{Ort}(H)^{\boxtimes n}, p^{\otimes n}) \leqslant 1$.*
(b) $p \in \mathscr{CE}^\infty(H)$ if and only if $\Theta(\mathrm{Ort}(H), p) \leqslant 1$, or, equivalently, if $\alpha(\mathrm{Ort}(H), p) = \Theta(\mathrm{Ort}(H), p) = 1$.

PROOF. (a) By definition, $p \in \mathscr{CE}^n(H)$ if and only if $\alpha(\mathrm{Ort}(H^{\otimes n}), p^{\otimes n}) \leqslant 1$. The claim now follows from Lemma 4.2.2.

(b) The first statement holds by the definition of Θ (A.6). For the second statement, $p \in \mathscr{CE}^\infty(H)$ implies that $\Theta(\mathrm{Ort}(H), p) \leqslant 1$. But since $\alpha(\mathrm{Ort}(H), p) = 1$ due to $p \in \mathscr{CE}^1(H)$, we find $\Theta(\mathrm{Ort}(H), p) = 1 = \alpha(\mathrm{Ort}(H), p)$. The converse is clear. □

If follows from Corollary 5.2.3 that $\mathcal{Q}(H) \subseteq \mathscr{CE}^\infty(H)$.

LEMMA 7.3.3. *For every $k, n \in \mathbb{N}$, the following inclusions hold:*

$$\mathscr{CE}^\infty(H) \subseteq \ldots \subseteq \ldots \mathscr{CE}^n(H) \subseteq \ldots \subseteq \mathscr{CE}^1(H).$$

This should be seen in contrast to Remark A.1.3.

PROOF. We choose any $p \in \mathscr{CE}^1(H)$. Thanks to Corollary A.3.11, we know

$$\alpha(\mathrm{Ort}(H)^{\boxtimes n}, p^{\otimes n}) \geqslant \alpha(\mathrm{Ort}(H)^{\boxtimes (n-1)}, p^{\otimes (n-1)}) \cdot \alpha(\mathrm{Ort}(H), p).$$

Now since $\alpha(\mathrm{Ort}(H), p) = 1$, the sequence $(\alpha(\mathrm{Ort}(H)^{\boxtimes n}, p^{\otimes n}))_{n \in \mathbb{N}}$ is monotonically nondecreasing. The claim now follows from Lemma 7.3.2. □

7.4. Does Consistent Exclusivity characterize the quantum set? In [**FSA+12**], we considered $\mathscr{CE}^\infty(B_{n,k,m})$ for Bell scenarios $B_{n,k,m}$ and asked whether it coincides with $\mathcal{Q}(B_{n,k,m})$. We will answer this question now.

PROPOSITION 7.4.1 (Navascués). *For every H,*

$$\mathcal{Q}_1(H) \subseteq \mathscr{CE}^\infty(H). \tag{7.2}$$

This observation was first made by Miguel Navascués, before this whole formalism had been set up. We are now in a position to give an essentially trivial proof.

PROOF. Combine Propositions 6.3.2 and Lemma 7.3.2 with A.3.9. □

In particular, together with $\mathcal{Q}(H) \subseteq \mathcal{Q}_1(H)$, this gives another proof of $\mathcal{Q}(H) \subseteq \mathscr{CE}^\infty(H)$, even if an excessively more convoluted one. This completes our exposition of Figure 1.

COROLLARY 7.4.2. *In the CHSH scenario, the LO principle does not characterize quantum models: $\mathcal{Q}(B_{2,2,2}) \subsetneq \mathscr{CE}^\infty(B_{2,2,2})$.*

PROOF. From 7.4.1, since $\mathcal{Q}(B_{2,2,2}) \subsetneq \mathcal{Q}_1(B_{2,2,2})$ [**NPA08**]. □

THEOREM 7.4.3. *There are contextuality scenarios H for which $\mathcal{Q}_1(H) \subsetneq \mathscr{CE}^\infty(H)$.*

PROOF. Our Proposition 6.3.2 and Lemma 7.3.2 suggests that this is related to the existence of graphs G for which $\alpha(G) = \Theta(G) < \vartheta(G)$. And indeed, we will turn Haemers' example [**Hae81**] of this phenomenon into an example of a contextuality scenario J_n with a probabilistic model $p_J \in \mathscr{CE}^\infty(H)$ with $p_J \notin \mathcal{Q}_1(H)$.

Let $n \geqslant 12$ be an integer divisible by 4. Let J_n have vertices $V(J_n)$ being all 3-element subsets of $\{1, \ldots, n\}$. An edge of J_n is given in terms of a partition of $\{1, \ldots, n\}$ into 4-element subsets; a vertex (3-element subset) belongs to the edge if and only if it is contained in one of the subsets of the partition.

By construction, all $e \in E(J_n)$ have cardinality $|e| = n$, since every partition consists of $n/4$ subsets and each subset hosts 4 vertices. Therefore, assigning a weight of $\frac{1}{n}$ to each vertex defines a

probabilistic model p_J. Now the non-orthogonality graph $\mathrm{Ort}(H_J)$ consists of the 3-element subsets of $\{1,\ldots,n\}$ two of which are adjacent if and only if they have exactly one element in common. This is the graph that was considered by Haemers [**Hae81**], who showed that

$$\alpha(\mathrm{Ort}(H_J)) = \Theta(\mathrm{Ort}(H_J)) = n < \vartheta(\mathrm{Ort}(H_J)).$$

Since the probabilistic model p_J has constant weights $\frac{1}{n}$, this means that

$$\alpha(\mathrm{Ort}(H_J), p_J) = \Theta(\mathrm{Ort}(H_J), p_J) = 1 < \vartheta(\mathrm{Ort}(H_J), p_J),$$

and hence $p_J \in \mathscr{CE}^\infty(H_J)$, but $p_J \notin \mathcal{Q}_1(H_J)$. □

7.5. Convexity and activation of Consistent Exclusivity. Since $\mathscr{CE}^1$ is defined in terms of linear inequalities, it is obviously convex. But what about the convexity of $\mathscr{CE}^n$ for $n \geqslant 2$ and $\mathscr{CE}^\infty$? Also, what about the analog of Proposition 6.3.3 for the $\mathscr{CE}^*$ family of sets? We will now see that these questions are intimately related to some challenging open problems in combinatorics which we introduce in Appendix A. Again, all the following results specialize to results about LO^∞ in the Bell scenario case.

CONJECTURE 7.5.1. *For all contextuality scenarios H, H_A, H_B,*
(a) $\mathscr{CE}^\infty(H_A) \otimes \mathscr{CE}^\infty(H_B) \subseteq \mathscr{CE}^1(H_A \otimes H_B)$;
(b) $\mathscr{CE}^\infty(H_A) \otimes \mathscr{CE}^\infty(H_B) \subseteq \mathscr{CE}^\infty(H_A \otimes H_B)$;
(c) $\mathscr{CE}^\infty(H)$ is convex.

Conjecture (b) states that the **activation** of non-membership in $\mathscr{CE}^\infty$ is impossible: if it fails, then there are H_A and H_B with $p_A \in \mathscr{CE}^\infty(H_A)$ and $p_B \in \mathscr{CE}^\infty(H_B)$, but $p_A \otimes p_B \notin \mathscr{CE}^\infty(H_A \otimes H_B)$, which we interpret as saying that p_A and p_B activate each other. Thanks to Corollary 5.2.3, such an activation would show that $p_A \notin \mathcal{Q}(H_A)$ or $p_B \notin \mathcal{Q}(H_B)$.

THEOREM 7.5.2. *(a) Each of these conjectures follows from its counterpart in Conjecture A.2.1.*
(b) Conjectures (a) and (b) are equivalent and imply (c).

PROOF. (a) Combine Proposition 6.3.2 and Lemma 7.3.2 with Propositions A.4.4 and A.4.6.
(b) Conjecture (b) clearly implies (a). For the converse, suppose that we have $p_A \in \mathscr{CE}^\infty(H_A)$ and $p_B \in \mathscr{CE}^\infty(H_B)$ with $p_A \otimes p_B \notin \mathscr{CE}^\infty(H_A \otimes H_B)$. Then there exists some $n \in \mathbb{N}$ with $(p_A \otimes p_B)^{\otimes n} \notin \mathscr{CE}^1(H_A^{\otimes n} \otimes H_B^{\otimes n})$. This would mean that $p_A^{\otimes n} \in \mathscr{CE}^\infty(H_A^{\otimes n})$ and $p_B^{\otimes n} \in \mathscr{CE}^\infty(H_B^{\otimes n})$ was a counterexample to 7.5.1(a).

Concerning the implication from (a) to 7.5.1(c), we consider $p_1, p_2 \in \mathscr{CE}^\infty(H)$ and deduce $p_1^{\otimes k} \otimes p_2^{\otimes(n-k)} \in \mathscr{CE}^1(H^{\boxtimes n})$ from assumption (a). Due to convexity of $\mathscr{CE}^1(H^{\otimes n})$, this shows that for any $\lambda \in (0,1)$,

$$(\lambda p_1 + (1-\lambda)p_2)^{\otimes n} = \sum_{k=0}^{n} \binom{n}{k} \lambda^k (1-\lambda)^{n-k} p_1^{\otimes k} \otimes p_2^{\otimes(n-k)} \in \mathscr{CE}^1(H^{\otimes n}),$$

so that $(\lambda p_1 + (1-\lambda)p_2) \in \mathscr{CE}^n(H)$. Since n was arbitrary, this means $(\lambda p_1 + (1-\lambda)p_2) \in \mathscr{CE}^\infty(H)$, as was to be shown. □

While we suspect each of the Conjectures 7.5.1 to actually be *equivalent* to its counterpart in A.2.1 (and in A.4.2), we have not been able to prove this yet.

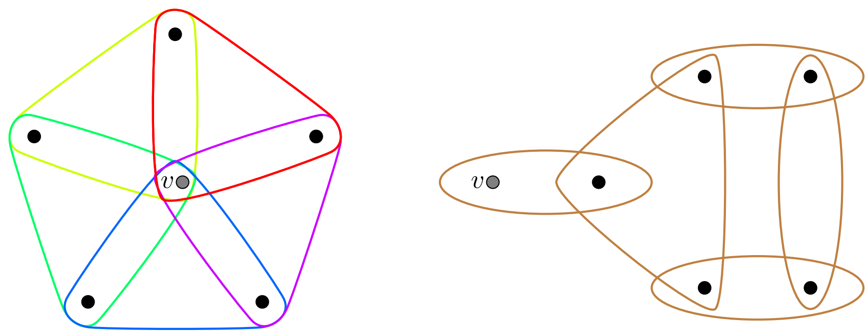


FIGURE 8. A scenario H_0 with $\mathcal{G}(H_0) = \mathcal{C}(H_0)$, although $\mathrm{Ort}(H_0)$ is not perfect. The two nodes labelled v represent the same vertex.

Failure of (b) or (c) would lead to an obvious way to strengthen the CE principle: the collection of physically realizable probabilistic models should be both convex and closed under $\otimes$. Therefore, if some physically realistic $q \in \mathscr{CE}^\infty(H)$ can be combined with some $p \in \mathscr{CE}^\infty(H)$ by using convex combinations and $\otimes$-products such that the combination is not in $\mathscr{CE}^\infty$, then p itself should be considered to violate the CE principle in a certain extended form.

7.6. Contextuality and perfection. We now study under which conditions $\mathcal{C}(H)$ coincides with $\mathscr{CE}^1(H)$. Recall that a graph G is called **perfect** if the chromatic number of any induced subgraph is equal to the clique number of this subgraph [**Ber61**].

PROPOSITION 7.6.1. *If* $\mathrm{Ort}(H)$ *is perfect, then* $\mathcal{C}(H) = \mathscr{CE}^1(H)$*, although* $\mathcal{G}(H)$ *can still be bigger.*

PROOF. By the weak perfect graph theorem of Lovász [**Lov72**], we can as well assume the complement $\overline{\mathrm{Ort}(H)}$ to be perfect. A probabilistic model $p \in \mathscr{CE}^1(H)$ can be interpreted as vertex weights $p(v)$ for $v \in V(H)$ with $\sum_{v\in C} p(v) \leqslant 1$ for every clique C in $\overline{\mathrm{Ort}(H)}$. Then, perfection guarantees [**Knu94**, Thm. 31] that p is a convex combination of indicator functions of independent sets in $\overline{\mathrm{Ort}(H)}$, i.e. there are cliques $U_1, \ldots, U_k$ in $\mathrm{Ort}(H)$ and coefficients $\lambda_i \in [0,1]$ with $\sum_i \lambda_i = 1$ such that

$$p = \sum_{i=1}^{k} \lambda_i \mathbb{1}_{U_i}. \tag{7.3}$$

We now claim that every $\mathbb{1}_{U_i}$ is a deterministic model. Since its weights clearly take values in $\{0,1\}$, it is enough to verify the normalization condition $\sum_{v\in e} \mathbb{1}_{U_i}(v) = 1$ for all $e \in E(H)$. But this follows from (7.3) together with $\sum_{v\in e} p(v)$.

In order to see that $\mathcal{G}(H)$ can still be bigger, consider again the triangle scenario Δ depicted in Figure 3. There, $\mathcal{C}(\Delta) = \mathscr{CE}^1(\Delta) = \varnothing$, although Δ allows a probabilistic model. □

The converse to Proposition 7.6.1 is not true:

PROPOSITION 7.6.2. *For the scenario depicted in Figure 8,* $\mathcal{G}(H_0) = \mathcal{C}(H_0)$*. However,* $\mathrm{Ort}(H_0)$ *is not perfect.*

PROOF. $\mathrm{Ort}(H_0)$ is not perfect since its complement $\overline{\mathrm{Ort}(H_0)}$ contains the pentagon $\pentagon$ as an induced subgraph in the left part.

On the other hand, every probabilistic model p on H_0 is guaranteed to satisfy $p(v) = 1$ due to the structure on the right. Hence, $p(u) = 0$ for all u in the pentagon. Therefore, both $\mathcal{G}(H_0)$ and $\mathcal{C}(H_0)$ can be identified with their counterparts for the right part H_R of Figure 8. Since every maximal independent set in $\mathrm{Ort}(H_R)$ is itself an edge, we get $\mathscr{CE}^1(H_R) = \mathcal{G}(H_R)$, and since $\mathrm{Ort}(H_R)$ is perfect, we have $\mathcal{C}(H_R) = \mathscr{CE}^1(H_R)$. □

Forcing the vanishing of the weights in the pentagon may seem like a cheap trick. However, we don't know of any natural combinatorial condition which one could impose on a contextuality scenario in order to exclude such pathological behavior of $\mathcal{G}(H)$. In particular, the proof of Shultz's Theorem 2.3.5 uses similar "forcing" ideas [**Shu74**].

See Proposition 9.3.3 for a slightly less artificial example of a scenario AP_4 with $\mathcal{Q}_1(\mathrm{AP}_4) = \mathscr{CE}^1(\mathrm{AP}_4)$, although $\mathrm{Ort}(\mathrm{AP}_4)$ is not perfect.

Theorem 7.6.3 (Strong perfect graph theorem [**CRST06**]). *A graph G is perfect if and only if neither G nor $\overline{G}$ contains an induced subgraph which is a cycle of odd length $\geqslant 5$.*

In combination together with Proposition 7.6.1, we obtain:

Corollary 7.6.4. *If neither $\mathrm{Ort}(H)$ nor $\overline{\mathrm{Ort}(H)}$ contains an odd cycle of length $\geqslant 5$ as an induced subgraph, then $\mathcal{C}(H) = \mathscr{CE}^1(H)$.*

In this sense, every (quantum) contextuality proof must rely on a "cycle-like" contradiction as it appears in the Klyachko-Can-Binicioğlu-Shumovsky scenario (see [**KCBS08**] and Section 9.2), or on an "anti-cycle-like" contradiction. Within the framework of [**CSW10**], this observation is due to [**CDLTP12**], where the anti-cycle case has been studied in a bit more detail.

8. Complexity of various decision problems

We now study the computational complexity of various decision problems associated to contextuality scenarios.

8.1. Deciding existence of probabilistic/classical models. The most basic decision problem about contextuality scenarios is this:

Problem: `ALLOWS_GENERAL`
Input: A contextuality scenario H,
Output: $\mathcal{G}(H) \neq \emptyset$?

Recall that there are indeed contextuality scenarios without any probabilistic models, for example Figure 4. Determining the complexity of `ALLOWS_GENERAL` is very basic as well:

Proposition 8.1.1. `ALLOWS_GENERAL` *is in* **P**.

Proof. Determining whether $\mathcal{G}(H) \neq \emptyset$ is a linear program. □

Now on to the analogous question about classical models:

Problem: `ALLOWS_CLASSICAL`
Input: A contextuality scenario H,
Output: $\mathcal{C}(H) \neq \emptyset$?

Proposition 8.1.2. `ALLOWS_CLASSICAL` *is* **NP**-*complete.*

PROOF. ALLOWS_CLASSICAL can be identified with the class of Boolean satisfiability problems which are disjunctions of clauses, where each clause states that exactly one variable in a certain subset of all variables needs to have the value TRUE. Given this, **NP**-completeness follows from Schaefer's dichotomy theorem [**Sch78**]. Notwithstanding, we now offer an explicit proof.

First, ALLOWS_CLASSICAL is clearly in **NP**: any explicit deterministic model $p : V(H) \to \{0,1\}$ witnesses $\mathcal{C}(H) \neq \varnothing$.

Let $x_1, \ldots, x_n$ be Boolean variables and

$$B = (y_{11} \vee y_{12} \vee y_{13}) \wedge \ldots \wedge (y_{m1} \vee y_{m2} \vee y_{m3}) \tag{8.1}$$

be a logical formula in which each y_{ij} stands for some x_l or its negation $\neg x_l$. The Boolean satisfiability problem 3SAT is the following decision problem:

Problem: 3SAT
Input: a logical formula B in the form (8.1),
Output: Is B satisfiable?

This is known to be **NP**-complete [**Kar72**]. We now prove **NP**-hardness of ALLOWS_CLASSICAL by polynomially reducing 3SAT to ALLOWS_CLASSICAL. Denote the clauses in B by

$$C_i = y_{i1} \vee y_{i2} \vee y_{i3}$$

and construct a contextuality scenario H_B as follows. We would like the set of vertices to correspond to the set of literals together with 8 auxiliary variables for each clause in the sense that

$$V(H_B) \stackrel{def}{=} \{v_{x_1}, \ldots v_{x_n}, v_{\neg x_1}, \ldots, v_{\neg x_n}\} \cup \{v_{i,s}\},$$

where $i = 1, \ldots, m$ enumerates the clauses and $s \in \{001, 010, 011, 100, 101, 110, 111\}$ runs over the feasible truth value assignments to the literals in a clause. There are three kinds of edges,

$$\begin{aligned} E(H_B) \stackrel{def}{=} & \left\{\{v_{x_j}, v_{\neg x_j}\} \;:\; j = 1, \ldots, n\right\} \cup \left\{\{v_{i,001}, \ldots, v_{i,111}\} \;:\; i = 1, \ldots, m\right\} \\ & \cup \left\{\{v_{i,s}, v_{(\neg)y_{i1}}, v_{(\neg)y_{i2}}, v_{(\neg)y_{i3}}\} \;:\; i = 1, \ldots, m;\; s = 001, \ldots, 111\right\} \end{aligned}$$

where in the third type of edge, the negation $\neg$ appears if and only if s has a 1 at the corresponding position.

The first type of edge guarantees that in any deterministic model, either v_{x_j} or $v_{\neg x_j}$ gets the value 1, but not both; the second kind of edge guarantees that for every i, exactly one of the $v_{i,s}$'s is 1; the third type ensures that if $p(v_{i,s}) = 1$, then the variables of C_i have precisely the values given by s. Therefore, the deterministic models on H_B correspond bijectively to the satisfying variable assignments of B. □

8.2. A semidefinite hierarchy converging to $\mathcal{C}(H)$. For some combinatorial optimization problems, one can find a contextuality scenario H, with polynomially many vertices and edges, such that the associated $\mathcal{C}(H)$ coincides with the usual polytope associated to the combinatorial optimization problem [**Sch03**]. We have illustrated how to do this for the case of 3SAT above, but similar reductions can be made also e.g. for coloring problems on graphs. The main idea is that the vertices of H are interpreted as boolean variables, and any formula of propositional logic can be encoded in terms of a collection of edges, possibly using some auxiliary variables. Then, our machinery automatically produces an associated linear as well a semidefinite relaxation of $\mathcal{C}(H)$, namely $\mathcal{G}(H)$ and $\mathcal{Q}_1(H)$, respectively. In fact, using a variant of Definition 6.1.2, where one additionally imposes that $M_{\mathbf{v},\mathbf{w}} = M_{\pi(\mathbf{v}),\mathbf{w}}$ for any permutation π, gives a **hierarchy of semidefinite relaxations** $(\mathcal{C}_n(H))_{n\in\mathbb{N}}$ converging to $\mathcal{C}(H)$ [**Las02**]. Due to the high number of constraints, this converges even after a finite number of steps: $\mathcal{C}_{|V(H)|}(H) = \mathcal{C}(H)$ since any matrix

element $M_{\mathbf{v},\mathbf{w}}$ with $\mathbf{v}$ or $\mathbf{w}$ longer than $V(H)$ is already determined by the other ones. However, while every $\mathcal{C}_n$ for fixed n is defined by a semidefinite program of polynomial size, the semidefinite program defining $\mathcal{C}_{|V|}$ is of exponential size. We have not implemented any of this since (hierarchies of) smaller specialized semidefinite relaxations exist in the literature [**Anj04**, **Lau03**].

8.3. Towards an inverse sandwich theorem? Now that we know the complexity of `ALLOWS_GENERAL` and `ALLOWS_CLASSICAL`, we move on to consider the quantum case, which may have some surprises to offer.

Problem: `ALLOWS_QUANTUM`
Input: A contextuality scenario H,
Output: $\mathcal{Q}(H) \neq \varnothing$?

This is equivalent to asking whether there exists an assignment of projections $P_v \in \mathcal{B}(\mathcal{H})$ to each $v \in V(H)$ such that $\sum_{v\in e} P_v = \mathbb{1}$ for all $e \in E(H)$. Here, the Hilbert space $\mathcal{H}$ can be taken to be separable infinite-dimensional without loss of generality, i.e. $\mathcal{H} = \ell^2(\mathbb{N})$.

By definition, every set $\mathcal{Q}_n(H)$ is given by a semidefinite program of polynomial size, and therefore determining whether $\mathcal{Q}_n(H) \neq \varnothing$ can be done efficiently. One might suspect that this should give an algorithm for `ALLOWS_QUANTUM` thanks to the following observation:

Lemma 8.3.1. *$\mathcal{Q}(H) = \varnothing$ if and only if $\mathcal{Q}_n(H) = \varnothing$ for some $n \in \mathbb{N}$.*

Proof. If $\mathcal{Q}_n(H) = \varnothing$ for some n, then clearly $\mathcal{Q}(H) = \varnothing$ as well. To show the converse, assume $\mathcal{Q}(H) = \varnothing$, so that $\bigcap_n \mathcal{Q}_n(H) = \varnothing$. Since this is an intersection of closed subsets of the compact space $\mathcal{Q}_1(H)$, we conclude that already finitely many of the $\mathcal{Q}_n(H)$ have empty intersection. Because the $\mathcal{Q}_n(H)$ form a decreasing sequence of sets, there has to be some $n \in \mathbb{N}$ with $\mathcal{Q}_n(H) = \varnothing$. □

However, simply checking whether $\mathcal{Q}_n(H) = \varnothing$ for some n is a procedure that never terminates if $\mathcal{Q}(H) \neq \varnothing$. Hence, in order to find an algorithm for `ALLOWS_QUANTUM`, we also need a procedure for detecting that $\mathcal{Q}(H) \neq \varnothing$ if this is the case.

One way to do this is to look in every Hilbert space dimension $\mathcal{H} = \mathbb{C}^d$ and see if there exists a quantum model in this dimension. For fixed d, this boils down to determining whether a certain system of polynomial equations and inequalities has a solution in $\mathbb{R}$. Thanks to real quantifier elimination [**Tar51**], this is known to be decidable. Therefore, if a quantum model over some finite-dimensional Hilbert space exists, this procedure will eventually find it.

By combining these two procedures, we have an algorithm for deciding `ALLOWS_QUANTUM` that works in all cases except in the case that H allows quantum models, but only on infinite-dimensional Hilbert spaces.

Problem 8.3.2. Does such an H exist?

In terms of [**FNT12**], we can also phrase this as follows. We construct the universal unital C^*-algebra associated to a contextuality scenario H,

$$C^*(H) \stackrel{def}{=} \left\langle \{P_v \;:\; v \in V(H)\} \;\middle|\; P_v = P_v^2 = P_v^* \;\forall v \in V(H), \quad \sum_{v\in e} P_v = \mathbb{1} \;\forall e \in E(H) \right\rangle$$

If all of these C^*-algebras are residually finite-dimensional, then Problem 8.3.2 has a negative answer and the above algorithm solves `ALLOWS_QUANTUM`, even if with very high complexity.

However, since Kirchberg's QWEP conjecture and Connes' embedding problem are equivalent to the residual finite-dimensionality of e.g. $C^*(B_{2,3,2})$ [**Fri12**, **FNT12**], we suspect that it is too much to hope for that all $C^*(H)$ are residually finite-dimensional.

A different approach to `ALLOWS_QUANTUM` lies in recognizing that any instance of it can be reformulated as an $\exists_1$ formula in quantum logic with signature $(\vee, \perp, \mathbb{1}_{\mathcal{H}})$ on an infinite-dimensional Hilbert space. However, since the decidability status of quantum logic is also not known [**Svo93**, p.69], this approach does not produce a terminating algorithm either.

In conclusion, we do not know of *any* terminating algorithm that would solve `ALLOWS_QUANTUM`. In fact, we suspect the following:

Conjecture 8.3.3. `ALLOWS_QUANTUM` *is undecidable.*

Recall that Lovász's **sandwich theorem** [**Knu94**] consists of the inequality

$$\alpha(G) \leqslant \vartheta(G) \leqslant \chi(\overline{G}) \tag{8.2}$$

together with the observation that the outer two quantities are **NP**-hard to compute, while $\vartheta(G)$ can be computed in polynomial time to arbitrary precision. We call 8.3.3 the **inverse sandwich conjecture** since the hypothetically uncomputable $\mathcal{Q}(H)$ lies between two computable sets,

$$\mathcal{C}(H) \subseteq \mathcal{Q}(H) \subseteq \mathcal{G}(H),$$

So in contrast to the case of (8.2), here the real meat indeed lies in the middle of the sandwich.

For the reasons discussed above, a proof of Conjecture 8.3.3 would have some interesting consequences for C^*-algebra theory and also prove the undecidability of quantum logic[3]. Since these are difficult problems in themselves, proving Conjecture 8.3.3 will also be difficult.

8.4. Other decision problems. There is a myriad of decision problems associated to contextuality scenarios which one can study. Some further ones are:

Problem: `IS_CLASSICAL`
Input: A contextuality scenario H and $p \in \mathcal{G}(H)$ with $p(v) \in \mathbb{Q}$,
Output: $p \in \mathcal{C}(H)$?

It is not difficult to see that this is in **NP**. Furthermore, it is actually **NP**-complete, since this is the case already for Bell scenarios [**AII06**].

Similarly, one can consider decision problems like `IS_QUANTUM` and `IS_LO`. So far, we have not considered these any further. Another natural decision problem is the question whether a given scenario allows nonclassical models or not:

Problem: `NONCONTEXTUAL`
Input: A contextuality scenario H,
Output: $\mathcal{C}(H) = \mathcal{G}(H)$?

We also do not know what the complexity of this problem is. We suspect that Theorem 2.4.3 together with the techniques of [**Eit94**] might be helpful for answering this question.

9. Examples

In the previous sections, we have developed the abstract theory of contextuality scenarios in some detail. We have exemplified some of the concepts and results for the case of Bell scenarios. In particular, this illustrates how our formalism makes precise the intuition that nonlocality is a

[3]More precisely, it would imply that the theory of Hilbert lattices in the signature $(\vee, \perp, \mathbb{1})$ is not decidable.

special case of contextuality. Also, Appendix B provides a large class of examples of contextuality scenarios.

Now it is time to look into other more concrete cases. The examples that have already been considered in the quantum foundations literature are too numerous to list. We focus on a few particularly appealing classes.

9.1. Hypergraphs with subnormalization. Cabello, Severini and Winter [**CSW10**] base their approach also on hypergraphs H in a very similar spirit as we have done. The main difference is that they do not impose a normalization constraint

$$\sum_{v\in e} p(v) = 1 \quad \forall e \in E(H),$$

but rather a subnormalization constraint

$$\sum_{v\in e} p(v) \leqslant 1 \quad \forall e \in E(H),$$

and similarly $\sum_{v\in e} P_v \leqslant \mathbb{1}_{\mathcal{H}}$ for projections in quantum representations. Our approach can "simulate" this behavior by constructing a contextuality scenario H' which contains one additional **no-detection event** w_e for each $e \in E(H)$,

$$V(H') \stackrel{def}{=} V(H) \cup \{w_e \,:\, e \in E(H)\}, \qquad E(H') \stackrel{def}{=} \{e \cup \{w_e\} \,:\, e \in E(H)\}\,.$$

Essentially by definition, the "classical models" of [**CSW10**] correspond to our $\mathcal{C}(H')$, the "generalized models" to our $\mathscr{CE}^1(H')$, and the "quantum models" to our $\mathcal{Q}_1(H')$.

Lemma 9.1.1. *In this situation,* $\mathcal{Q}(H') = \mathcal{Q}_1(H')$*.*

Proof. Starting from $p \in \mathcal{Q}_1(H')$, we would like to show that $p \in \mathcal{Q}(H')$. We use 6.3.1(d) as a criterion for membership in $\mathcal{Q}_1(H')$. This means that we have projections P_v for all $v \in V(H)$ and P_{w_e} for all $e \in E(H)$ such that

$$u \perp v \implies P_u \perp P_v, \qquad v \in e \implies P_v \perp P_{w_e},$$

and $p(v) = \langle\Psi|P_v|\Psi\rangle$ as well as $p(w_e) = \langle\Psi|P_{w_e}|\Psi\rangle$. We now define

$$P'_{w_e} \stackrel{def}{=} \mathbb{1}_{\mathcal{H}} - \sum_{v\in e} P_v$$

and claim that these, together with the P_v and the state $|\Psi\rangle$, form a quantum model of p. First, due to $\sum_{v\in e} P_v \leqslant \mathbb{1}_{\mathcal{H}}$, the operator P'_{w_e} is also a projection. Second, the completeness relation for edges in $E(H')$ then holds by definition. Third,

$$\langle\Psi|P'_{w_e}|\Psi\rangle = \langle\Psi|\Psi\rangle - \sum_{v\in e}\langle\Psi|P_v|\Psi\rangle = 1 - \sum_{v\in e} p(v) = p(w_e).$$

as claimed. Hence, $p \in \mathcal{Q}(H')$. □

In this sense, the set of quantum models of a scenario which arises in this way is particularly simple: the whole semidefinite hierarchy collapses to the first level! The advantage of this is that Proposition 6.3.2 on the relation between $\mathcal{Q}_1$ and the Lovász number now also applies to quantum models. So, scenarios constructed in this way form a very special and well-behaved subclass of all contextuality scenarios. The n-circular hypergraphs that we consider next arise in this way. However, many of the more interesting contextuality scenarios—like Bell scenarios—are not of this form and therefore cannot be treated correctly in the CSW approach. This was already noticed in [**CSW10**].

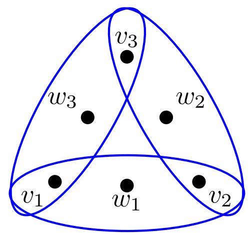


FIGURE 9. The 3-circular hypergraph Δ_3. The labeling of the vertices corresponds to [**FGR92**, Ex. 2.13].

9.2. n-circular hypergraphs. The n-circular hypergraphs generalize the "pentagon" idea of Klyachko-Can-Binicioğlu-Shumovsky [**KCBS08**].

DEFINITION 9.2.1. *For $n \geqslant 3$, the **n-circular hypergraph** Δ_n is given by*

$$V(\Delta_n) \stackrel{def}{=} \{v_1, \ldots, v_n, w_1, \ldots, w_n\},$$
$$E(\Delta_n) \stackrel{def}{=} \{\{v_1, w_1, v_2\}, \ldots, \{v_n, w_n, v_1\}\}.$$

So, Δ_n has $2n$ vertices and n edges as follows: if all vertices are evenly distributed on a circle in the order $v_1, w_1, \ldots, v_n, w_n, v_1$, then every second triple of adjacent vertices, namely those of the form $\{v_j, w_j, v_{j+1}\}$, is an edge (we write $v_{n+1} = v_1$). The w_i can be interpreted as no-detection events as explained in the previous subsection. In particular, Lemma 9.1.1 applies, and we see that $\mathcal{Q}(\Delta_n) = \mathcal{Q}_1(\Delta_n)$.

Figure 9 displays Δ_3, which can be metaphorically illustrated as a **firefly box** [**Wil08**]. It corresponds to the **Wright triangle** of [**FGR92**, Ex. 2.13] under the relabeling

$$v_1 \mapsto a, \quad w_1 \mapsto b, \quad v_2 \mapsto c, \quad w_2 \mapsto d, \quad v_3 \mapsto e, \quad w_3 \mapsto f.$$

Δ_5 is the "pentagon" scenario on which the KCBS inequality [**KCBS08**] is defined. It was first considered by Wright in 1978 [**Wri78**]. We now extend some of these results to arbitrary n.

PROPOSITION 9.2.2. *Let $n \geqslant 3$.*

(a) $\dim(\mathcal{C}(\Delta_n)) = \dim(\mathcal{G}(\Delta_n)) = n$.
(b) If n is even, then $\mathcal{C}(\Delta_n) = \mathcal{G}(\Delta_n)$.
(c) If n is odd, then $\mathcal{C}(\Delta_n) \subsetneq \mathcal{G}(\Delta_n)$ is determined by the inequality

$$\sum_i p(v_i) \leqslant \frac{n-1}{2}. \tag{9.1}$$

There is one extreme point of $\mathcal{G}(\Delta_n)$ which violates this inequality. It is the probabilistic model $p_x \in \mathcal{G}(\Delta_n)$ with

$$p_x(v_i) = \tfrac{1}{2}\ \forall i, \qquad p_x(w_i) = 0\ \forall i. \tag{9.2}$$

In particular, $\mathcal{G}(\Delta_n)$ has one vertex more than $\mathcal{C}(\Delta_n)$.

PROOF. We consider all vertex indices modulo n, so that $v_{n+1} = v_1$ etc.

(a) The equations imposed on the probabilities $p(v_i)$ and $p(w_i)$ by the normalization constraints are just

$$p(w_i) = 1 - p(v_i) - p(v_{i+1}), \tag{9.3}$$

which implies $\dim(\mathcal{G}(\Delta_n)) \leqslant n$. The conclusion follows if we can produce $n+1$ linearly independent deterministic models. This is simple: the set of models

$$p_j(v_i) \stackrel{def}{=} \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise,} \end{cases}$$

where the $p_j(w_i)$ are uniquely determined thanks to (9.3) and $j \in \{1, \ldots, n\}$, is linearly independent. Furthermore, adding to this set the model p_0 with $p_0(v_i) = 0$ for all i preserves linear independence. This is the desired collection of $n+1$ linearly independent deterministic models.

(b) $\mathcal{C}(\Delta_n) = \mathscr{CE}^1(\Delta_n)$ follows from Corollary 7.6.4, and $\mathscr{CE}^1(\Delta_n) = \mathcal{G}(\Delta_n)$ because the maximal independent sets of $\mathrm{Ort}(\Delta_n)$ are precisely the edges on Δ_n. In particular, while (9.2) is also a probabilistic model for even n, in this case it has to be a convex combination of deterministic models.

Note that this argument has not used (a).

(c) We apply Theorem 2.4.3 in combination with Corollary 7.6.4. Any induced subscenario H_W with $\mathcal{C}(H_W) \neq \mathcal{G}(H_W)$ needs to contain an induced (anti-)cycle in $\mathrm{Ort}(H_W)$. This is possible only if W contains all v_i. If W also contains one or more of the w_i's, then H_W does not have a unique probabilistic model. Therefore, there can be at most one nonclassical extreme point of $\mathcal{G}(H)$, namely the one associated to the induced subscenario on $W = \{v_1, \ldots, v_n\}$. Now this H_W does indeed have a unique probabilistic model given by $p_x(v_i) = \frac{1}{2}$, which yields (9.2) upon extension to Δ_n. This proves that $\mathcal{G}(\Delta_n)$ has p_x as its sole nonclassical extreme point without ever using any inequalities.

We now give an independent proof showing that (9.1) defines $\mathcal{C}(\Delta_n)$. Thanks to (9.3), it is enough to consider the values $p(v_i)$ only. Now the deterministic models correspond to the independent sets in the cycle graph C_n; upon identifying each vertex with the edge adjacent on its left, an independent set in C_n gets identified with a set of edges in C_n no two of which are adjacent at the same vertex, i.e. with a **matching** on C_n. Now it is known [**Sch03**] that the polytope of all matchings on C_n corresponds to

$$p(v_i) \geqslant 0, \qquad p(v_i) + p(v_{i+1}) \leqslant 1, \qquad \sum_{i=1}^{n} p(v_i) \leqslant \frac{n-1}{2}.$$

This is precisely the description of $\mathcal{C}(\Delta_n)$ that was to be proven.

□

For $n = 5$, the set of classical models is bounded by the inequality $\sum_{i=1}^{5} p(v_i) \leqslant 2$, which is precisely the inequality which has been studied in [**KCBS08**]. Compare [**AQB⁺12**] for the characterization of classical models in a related scenario.

PROPOSITION 9.2.3. $\mathcal{C}(\Delta_3) = \mathscr{CE}^1(\Delta_3) \subsetneq \mathcal{G}(\Delta_3)$. *For all other n, $\mathscr{CE}^1(\Delta_n) = \mathcal{G}(\Delta_n)$.*

PROOF. Since $\{v_1, v_2, v_3\}$ is the only independent set in $\mathrm{Ort}(\Delta_3)$ which is not an edge of Δ_3, we find that $\mathscr{CE}^1(\Delta_3)$ as a subset of $\mathcal{G}(\Delta_3)$ is given by imposing the inequality $p(v_1)+p(v_2)+p(v_3) \leqslant 1$. This is precisely the inequality that determines $\mathcal{C}(\Delta_3)$ in 9.2.2(c). For $n \geqslant 4$, however, every independent set in $\mathrm{Ort}(\Delta_n)$ is of the form $\{v_i, w_i, v_{i+1}\}$, i.e. is itself an edge. □

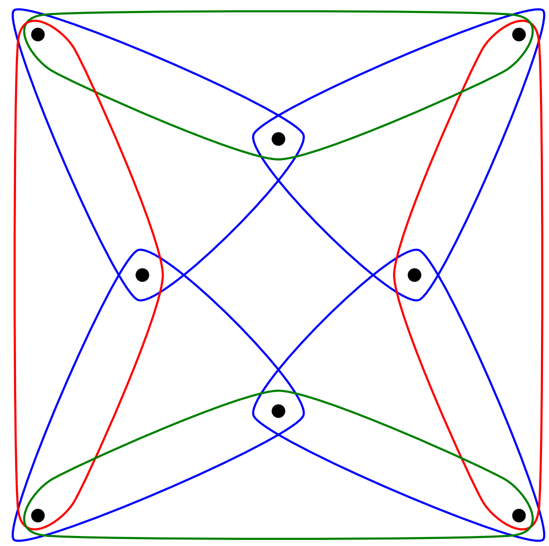


FIGURE 10. The contextuality scenario AP_4.

9.3. Antiprism scenarios. The antiprism scenarios are a variant of the circular hypergraph scenarios with some additional edges thrown in such that there is a symmetry exchanging the v_i with the w_i. Again, we consider all vertex indices modulo n. The antiprism scenarios are supposed to illustrate that an interesting looking hypergraph is not necessarily an interesting contextuality scenario.

DEFINITION 9.3.1. *Let $n \geqslant 3$. The n-**antiprism scenario** AP_n is*

$$\begin{aligned} V(AP_n) &\stackrel{def}{=} \{v_1, \ldots, v_n, w_1, \ldots, w_n\}, \\ E(AP_n) &\stackrel{def}{=} \{\{v_1, w_1, v_2\}, \ldots, \{v_n, w_n, v_1\}\} \\ &\qquad \cup \{\{w_1, v_2, w_2\}, \ldots, \{w_n, v_1, w_1\}\}. \end{aligned}$$

The idea behind the term "antiprism" is that one gets AP_n by considering the antiprism polytope over an n-gon and defines a hypergraph AP_n as given by the band of triangles winding itself around the polytope.

PROPOSITION 9.3.2. *If n is divisible by 3, then $\mathcal{C}(AP_n) = \mathcal{G}(AP_n)$ is a 2-dimensional triangle. Otherwise, AP_n has a unique probabilistic model which is not classical.*

PROOF. We show that $p(v_1)$ and $p(v_2)$ determine all other probabilities $p(v_i)$ and $p(w_i)$ by induction on i:

$$p(v_{i+1}) = 1 - p(v_i) - p(w_i), \qquad p(w_{i+1}) = 1 - p(w_i) - p(v_{i+1}).$$

In fact, this shows that

$$p(v_{3j+1}) = p(w_{3j+2}) = p(v_1), \quad p(v_{3j+2}) = p(w_{3j}) = p(v_2), \quad p(v_{3j}) = p(w_{3j+1}) = 1 - p(v_1) - p(v_2).$$

Now if n is divisible by 3, then this is consistent upon "going around the cycle", so that $\mathcal{G}(AP_n)$ can be identified with the triangle

$$p(v_1) \geqslant 0, \qquad p(v_2) \geqslant 0, \qquad p(v_1) + p(v_2) \leqslant 1.$$

Clearly, the extreme points of this triangle are deterministic.

If n is not divisible by 3, then the above recurrence relations imply that $p(v_1) = p(v_2) = \frac{1}{3}$, so that $\mathcal{G}(AP_n)$ degenerates to a single point. $\mathcal{C}(AP_n) = \varnothing$ since there is no deterministic model. $\square$

We now give another example application of our methods.

PROPOSITION 9.3.3. *$\mathcal{Q}_1(AP_4) = \varnothing$, although $\mathscr{CE}^1(AP_4) = \mathcal{G}(AP_4)$.*

PROOF. Direct inspection shows that every maximal independent set in $\mathrm{Ort}(AP_4)$ is an edge, so that the unique probabilistic model given by $p(v_i) = p(w_i) = \frac{1}{3}$ is in $\mathscr{CE}^1(AP_4)$.

It remains to show that the unique probabilistic model is not in $\mathcal{Q}_1(AP_4)$. By Proposition 6.3.2, this boils down to showing that $\frac{1}{3}\vartheta(\mathrm{Ort}(AP_4)) > 1$. Now $\mathrm{Ort}(AP_n)$ is the complement of the 4-antiprism graph $\bowtie_4$. Since $\bowtie_4$ is vertex-symmetric, we deduce [**Knu94**, Thm. 25] that $\vartheta(\bowtie_4)\vartheta(\mathrm{Ort}(AP_4)) = 8$. Now $\vartheta(\bowtie_4)$ is known [**BPT11**] to equal $8 - 4\sqrt{2}$, so that

$$\vartheta(\mathrm{Ort}(AP_4)) = \frac{8}{8 - 4\sqrt{2}} = \frac{2}{2 - \sqrt{2}} = 2 + \sqrt{2} > 3,$$

as was to be shown. $\square$

Also, note that the antiprism graph $\bowtie_4$ which appears in this proof has also arisen as the non-orthogonality graph of possible events for the PR-box [**Cab12b**, **FSA$^+$12**].

9.4. Matching scenarios. Let K_m be the complete graph on m vertices. We define a contextuality scenario Mat_m as follows. $V(\mathrm{Mat}_m)$ is defined to be the set of *edges* of K_m, so that $|V(\mathrm{Mat}_m)| = \frac{m(m-1)}{2}$. The set of edges of Mat_m itself is $E(\mathrm{Mat}_m) = \{e_1, \ldots, e_m\}$, where e_j is the set of all edges in K_m adjacent to the vertex $j \in K_m$. In the language of hypergraph theory [**Vol09**], Mat_m is the **dual** of K_m. For reasons that will become clear, we call it a **matching scenario**.

Mat_5 coincides with Figure 2(b) from [**PMMM05**]. Using the CSW formalism [**CSW10**], it has also recently been studied in [**Cab12a**]. These latter results can be transferred to our setting using the construction of Section 9.1, but they will live in the contextuality scenario Mat'_5 which contains additional vertices representing no-detection events.

There are certain probabilistic models on $\mathrm{Mat_m}$ which have a special form. By a **half-integer matching**, we mean a probabilistic model on Mat_m in which each probability lies in $\{0, \frac{1}{2}, 1\}$ in such a way that the edges with positive probability define a decomposition of K_m into cycles of odd length, where we regard an edge of probability 1 as a cycle of length 1. In particular, every **perfect matching** on K_m can be regarded as a half-integer matching.

PROPOSITION 9.4.1. *(a) The deterministic models on Mat_m are the perfect matchings on K_m.*
*(b) $\mathcal{C}(\mathrm{Mat}_m)$ is the perfect matching polytope [**Sch03**] on K_m. In particular, $\mathcal{C}(\mathrm{Mat_m}) \neq \varnothing$ if and only if m is even.*
(c) $\mathcal{G}(\mathrm{Mat}_m)$ is the fractional matching polytope. Its extreme points are precisely the half-integer matchings.
(d) $\mathscr{CE}^1(\mathrm{Mat}_m)$ is a polytope strictly intermediate between $\mathcal{C}(\mathrm{Mat_m})$ and $\mathcal{G}(\mathrm{Mat_m})$ for $m \geqslant 5$.

Proof. (a) Using Remark 4.1, a deterministic model corresponds to a collection of edges in K_m such that there is exactly one edge incident to each vertex. This is the definition of perfect matching.

(b) $\mathcal{C}(\mathrm{Mat}_m)$ is defined to be the convex hull of the deterministic models, and likewise the perfect matching polytope is defined to be the convex hull of the perfect matchings, in the same ambient space.

(c) The inequalities defining $\mathcal{G}(\mathrm{Mat}_m)$ are precisely those defining the standard linear relaxation of the perfect matching polytope. Its extreme points are known to be the half-integer matchings [**Sch03**]. This can also be proven using Theorem 2.4.3.

(d) For $m \geqslant 5$, there are two kinds of maximal independent sets in $\mathrm{Ort}(\mathrm{Mat}_m)$: first, the edges of Mat_m themselves; second, all triples of edges in K_m that form a triangle. The latter impose the additional constraint that the sum of the edge weights in a triangle should not exceed 1. Therefore, the half-integer matchings with cycles of length 3 do not belong to $\mathscr{CE}^1(\mathrm{Mat}_m)$, which is hence a polytope strictly contained in $\mathcal{G}(\mathrm{Mat}_m)$. On the other hand, $\mathscr{CE}^1(\mathrm{Mat}_m)$ still contains half-integer matchings with odd cycles of length $\geqslant 5$, which are not in $\mathcal{C}(\mathrm{Mat}_m)$.

□

Nothing of this is specific to K_m and can likewise be done starting with any other graph.

By definition, $\mathrm{Ort}(\mathrm{Mat}_m)$ is the Kneser graph $KG_{m,2}$ [**Lov78**]. In particular, $\mathrm{Ort}(\mathrm{Mat}_5)$ is the Petersen graph. But the curiosities do not end here:

Corollary 9.4.2. $\mathscr{CE}^1(\mathrm{Mat}_5)$*, when scaled by a factor of* 2*, is the symmetric traveling salesman polytope* STSP(5) [**GP79**, **NP01**].

Proof. Since 5 is odd, K_5 has no perfect matchings. Therefore, every half-integer matching on K_m is a disjoint union of cycles of edges with weight $\frac{1}{2}$. Now it follows from (d) that every extremal vertex of $\mathscr{CE}^1(\mathrm{Mat}_m)$ is a cycle of length 5 with weight $\frac{1}{2}$ on each edge, as was to be shown. □

Appendix A. Background on graph theory

This section reviews standard material on the invariants of graphs which are of relevance to the main text, first for unweighted and then for weighted graphs. As far as we know, the conjectures in Sections A.2 and A.4 are new.

For us, a **graph** is an undirected simple graph without isolated vertices. When G is a graph, we denote its set of vertices by $V(G)$. For $u, v \in V(G)$, we write $u \sim_G v$ whenever u and v share an edge (are **adjacent**) in G. Usually the graph is clear, and then we simply write $u \sim v$.

There are many ways to take products of graphs [IK00]. For us, the relevant one is this:

Definition A.0.1. *Let G_1 and G_2 be graphs. Their* ***strong product*** *is the graph $G_1 \boxtimes G_1$ with*

$$V(G_1 \boxtimes G_2) \stackrel{def}{=} V(G_1) \times V(G_2)$$

and $(u_1, u_2) \sim (v_1, v_2)$ whenever

$$(u_1 \sim v_1 \wedge u_2 \sim v_2) \vee (u_1 \sim v_1 \wedge u_2 = v_2) \vee (u_1 = v_1 \wedge u_2 \sim v_2)\,.$$

For $n \in \mathbb{N}$, we write $G^{\boxtimes n}$ for the n-fold strong product of G with itself.

A.1. Relevant invariants of unweighted graphs.

Since we will later consider graphs equipped with edge weights, we also use the term "unweighted graph" when working with plain graphs in order to emphasize the distinction.

Recall that an **independent set** in a graph G is a subset $I \subseteq V(G)$ such that no two vertices in I share an edge. I is an independent set in G if and only if it is a **clique** in the complement graph $\overline{G}$. An independent set I is **maximal** if there is no other independent set $I' \subseteq V(G)$ with $I \subsetneq I'$. The **independence number** $\alpha(G)$ is the largest number of elements of any independent set in G (sometimes also called the **stability number**).

Lemma A.1.1. *Let $I_1 \subseteq G_1$ and $I_2 \subseteq G_2$ be maximal independent sets. Then $I_1 \times I_2 \subseteq G_1 \boxtimes G_2$ is also a maximal independent set.*

Proof. The definition of adjacency in $G_1 \boxtimes G_2$ implies immediately that $I_1 \times I_2$ is also an independent set in $G_1 \boxtimes G_2$.

We now show maximality of $I = I_1 \boxtimes I_2$. For any $v = (v_1, v_2) \in V(G_1 \boxtimes G_2) \backslash I$, the following cases are possible:

(a) Case $v_1 \notin I_1$ and $v_2 \notin I_2$: by maximality of I_1 and I_2, there are $u_1 \in I_1$ with $u_1 \sim v_1$ and $u_2 \in I_2$ with $u_2 \sim v_2$. Hence $(u_1, u_2) \sim (v_1, v_2)$.
(b) Case $v_1 \notin I_1$ and $v_2 \in I_2$: by maximality of I_1, there is $u_1 \in I_1$ with $u_1 \sim v_1$. Hence $(u_1, v_2) \in I$ and $(u_1, v_2) \sim (v_1, v_2)$.
(c) Case $v_1 \in I_1$ and $v_2 \notin I_2$: Similar to the previous case.

In either case, the conclusion is that v is adjacent to some vertex in I, and hence I is a maximal independent set. □

Lemma A.1.2.

$$\alpha(G_1 \boxtimes G_2) \geqslant \alpha(G_1)\alpha(G_2)$$

Proof. Lemma A.1.1. □

In particular, this implies

$$\alpha(G^{\boxtimes (n+m)}) \geqslant \alpha(G^{\boxtimes n})\alpha(G^{\boxtimes m}) \qquad \forall m, n \in \mathbb{N}. \tag{A.1}$$

REMARK A.1.3. Despite this inequality, the sequence $\left(\sqrt[n]{\alpha(G^{\boxtimes n})}\right)_{n\in\mathbb{N}}$ is not monotonically increasing in general; this happens, for example, for the pentagon graph (or 5-cycle) $\pentagon$, for which

$$\alpha(\pentagon) = 2, \qquad \alpha(\pentagon^{\boxtimes 2}) = 5, \qquad \alpha(\pentagon^{\boxtimes 3}) = 10.$$

See [**AL06**] for more results on the behavior of $\left(\sqrt[n]{\alpha(G^{\boxtimes n})}\right)_{n\in\mathbb{N}}$.

In combination with Fekete's Lemma [**Fek23**], (A.1) guarantees the existence of the following limit:

DEFINITION A.1.4 (Shannon capacity). *The* ***(unweighted) Shannon capacity*** $\Theta(G)$ *is*

$$\Theta(G) \stackrel{def}{=} \lim_{n\to\infty} \sqrt[n]{\alpha(G^{\boxtimes n})}. \tag{A.2}$$

Intuitively, $\Theta(G)$ is an asymptotic version of the independence number $\alpha(G)$.

This number can be interpreted in terms of information theory as follows. When $V(G)$ is the input alphabet of a classical communication channel such that $u \sim v$ if and only if u and v have non-trivial probability to produce the same channel output, then this channel can asymptotically transfer $\log_2 \Theta(G)$ bits of perfect information per channel use. G is then called the **confusability graph** of the channel. This is the context in which Θ was originally introduced by Shannon [**Sha56**]. The use of the logarithm here differs from the standard information-theoretic definitions of capacities, which usually already include it in their definition.

Not much is known about the values of Θ for particular graphs, not even $\Theta(C_7)$, where C_7 is the 7-cycle [**CGR03**].

For graphs G_1 and G_2, we write $G_1 + G_2$ for their disjoint union, which is again a graph.

LEMMA A.1.5 ([**Sha56**]). *(a)*

$$\Theta(G_1 + G_2) \geqslant \Theta(G_1) + \Theta(G_2). \tag{A.3}$$

(b)

$$\Theta(G_1 \boxtimes G_2) \geqslant \Theta(G_1)\Theta(G_2). \tag{A.4}$$

Finding examples in which these inequalities are not tight is surprisingly difficult. The following results are due to Haemers and Alon.

THEOREM A.1.6 ([**Hae79**, **Alo98**]). *There exist graphs G_1 and G_2 such that*

(a) $\Theta(G_1 \boxtimes G_2) > \Theta(G_1)\Theta(G_2)$.
(b) $\Theta(G_1 + G_2) > \Theta(G_1) + \Theta(G_2)$,

Of particular relevance for our considerations in the main text are graphs whose independence number coincides with their Shannon capacity:

DEFINITION A.1.7. *A graph G is* ***single-shot*** *if $\alpha(G) = \Theta(G)$.*

Single-shot graphs are the *Class 1* graphs of Berge [**Ber97**][4]. G is single-shot precisely when the sequence $\left(\sqrt[n]{\alpha(G^{\boxtimes n})}\right)_{n\in\mathbb{N}}$ is constant. Our terminology is motivated by the information-theoretic interpretation alluded to above: if a communication channel has a confusability graph which is single-shot, then there exists a zero-error code for this channel which operates on the single-shot level.

[4] We thank András Salamon for pointing out this reference.

Due to standard results [**Knu94**], every perfect graph is single-shot. The Petersen graph is not perfect, but nevertheless single-shot since its Lovász number (see below) coincides with its independence number [**Knu94**, p. 31].

Definition A.1.8 (Lovász number [**Lov79**]). *(a) An* ***orthonormal labeling*** *of G is an assignment $v \mapsto |\psi_v\rangle$ of a unit vector $|\psi_v\rangle \in \mathbb{R}^{|V(G)|}$ to every $v \in V(G)$ such that $u \not\sim v$ and $u \neq v$ implies $|\psi_u\rangle \perp |\psi_v\rangle$.*

(b) The ***Lovász number*** *$\vartheta(G)$ is*

$$\vartheta(G) \stackrel{def}{=} \min_{|\Psi\rangle,\, |\psi_v\rangle} \max_{v\in V} \frac{1}{|\langle\Psi|\psi_v\rangle|^2}$$

where $|\Psi\rangle \in \mathbb{R}^{|V(G)|}$ ranges over all unit vectors and $(|\psi_v\rangle)_{v\in V(G)}$ over all orthonormal labelings.

There are several other equivalent definitions of $\vartheta(G)$ commonly used [**Lov79**]. Multiplicativity of ϑ is one of its many useful properties:

Proposition A.1.9 ([**Lov79**]).

$$\vartheta(G_1 \boxtimes G_2) = \vartheta(G_1)\vartheta(G_2).$$

Definition A.1.10. *The* ***fractional packing number*** *$\alpha^*(G)$ is*

$$\alpha^*(G) \stackrel{def}{=} \max_q \sum_v q_v$$

where $q : V(G) \to [0,1]$ ranges over all vertex weighings satisfying $\sum_{v\in C} q_v \leqslant 1$ for all cliques $C \subseteq V(G)$.

The fractional packing number can be regarded as the linear relaxation of the independence number. For this reason, it is sometimes also called **fractional independence number**.

Proposition A.1.11 ([**Lov79**]).

$$\alpha(G) \leqslant \Theta(G) \leqslant \vartheta(G) \leqslant \alpha^*(G).$$

In general, none of these inequalities is an equality. This is most difficult to see for $\Theta(G) \leqslant \vartheta(G)$, for which it was shown by Haemers [**Hae79**] after having been posed as an open problem by Lovász [**Lov79**].

A.2. Main conjectures for unweighted graphs. Single-shot graphs are of particular relevance for the material in Section 7, and therefore we would like to understand some of their properties.

In the following, we routinely use the material of the previous subsection without explicit reference.

Conjecture A.2.1. *For all single-shot graphs G_1, G_2,*

(a) $\alpha(G_1 \boxtimes G_2) = \alpha(G_1)\alpha(G_2)$;
(b) $\Theta(G_1 \boxtimes G_2) = \Theta(G_1)\Theta(G_2)$;
(c) $G_1 + G_2$ is also single-shot, so that $\Theta(G_1 + G_2) = \Theta(G_1) + \Theta(G_2)$.

Concerning (c), we have

$$\alpha(G_1+G_2)=\alpha(G_1)+\alpha(G_2)=\Theta(G_1)+\Theta(G_2)\leqslant\Theta(G_1+G_2),$$

so that G_1+G_2 is also single-shot if and only if $\Theta(G_1+G_2)=\Theta(G_1)+\Theta(G_2)$. However, it is not clear whether (b) is similarly related to the question whether $G_1\boxtimes G_2$ is also single-shot.

We are far from being able to answer any of these conjectures. Using Proposition A.3.10, it is clear that (a) holds in the class of those graphs which satisfy $\alpha(G)=\Theta(G)=\vartheta(G)$, and (b) and (c) then follow from the upcoming Proposition A.2.2; however, it was shown by Haemers [**Hae81**] that there are graphs G with $\alpha(G)=\Theta(G)<\vartheta(G)$. Moreover, the other examples of Haemers [**Hae79**] or Alon [**Alo98**] seem to have no bearing on the above conjectures, although Haemers' bound might also turn out to be useful for finding counterexamples. We also hope that our Theorem 7.5.2 might be useful for finding counterexamples since it allows to ponder a large class of specific cases in terms of physical intuition.

Despite the difficulty of these conjectures, it is relatively simple to show some interrelations between them:

PROPOSITION A.2.2. *Conjectures (a) and (b) are equivalent and imply (c).*

PROOF. Assuming (b), we get

$$\alpha(G_1)\alpha(G_2)\leqslant\alpha(G_1\boxtimes G_2)\leqslant\Theta(G_1\boxtimes G_2)=\Theta(G_1)\Theta(G_2)=\alpha(G_1)\alpha(G_2),$$

which proves (a). Conversely, assuming (a) gives

$$\Theta(G_1\boxtimes G_2)=\lim_n\sqrt[n]{\alpha(G_1^{\boxtimes n}\boxtimes G_2^{\boxtimes n})}=\lim_n\sqrt[n]{\alpha(G_1^{\boxtimes n})\alpha(G_2^{\boxtimes n})}=\Theta(G_1)\Theta(G_2),$$

since if G_1 and G_2 are single-shot, then so are $G_1^{\boxtimes n}$ and $G_2^{\boxtimes n}$ for any $n\in\mathbb{N}$, and (a) then also applies to these graphs.

Concerning (c), we use the fact that $\boxtimes$ distributes over $+$, so that

$$(G_1+G_2)^{\boxtimes n}=\sum_{k=0}^{n}\binom{n}{k}G_1^{\boxtimes k}G_2^{\boxtimes(n-k)}.$$

This gives

$$\Theta(G_1+G_2)=\lim_n\sqrt[n]{\alpha\left(\sum_{k=0}^{n}\binom{n}{k}G_1^{\boxtimes k}\boxtimes G_2^{\boxtimes(n-k)}\right)}=\lim_n\sqrt[n]{\sum_{k=0}^{n}\binom{n}{k}\alpha\left(G_1^{\boxtimes k}\boxtimes G_2^{\boxtimes(n-k)}\right)}.$$

Applying (a) to the single-shot graphs $G_1^{\boxtimes k}$ and $G_2^{\boxtimes(n-k)}$ evaluates this to

$$\Theta(G_1+G_2)=\sqrt[n]{\sum_{k=0}^{n}\binom{n}{k}\alpha(G_1)^k\alpha(G_2)^{n-k}}=\alpha(G_1)+\alpha(G_2)=\Theta(G_1)+\Theta(G_2).$$

□

A.3. Relevant invariants of weighted graphs. We now generalize the definitions to graphs equipped with vertex weights, i.e. to graphs G equipped with a **weight function** $p:V(G)\to\mathbb{R}_+$. We omit a proof whenever it is completely analogous to the unweighted case. Weight functions $p_1:V(G_1)\to\mathbb{R}_+$ and $p_2:V(G_2)\to\mathbb{R}_+$ multiply to a weight function

$$p_1\otimes p_2:V(G_1\boxtimes G_2)\to\mathbb{R}_+,\quad (v_1,v_2)\mapsto p_1(v_1)p_2(v_2).$$

In this way, $p^{\otimes n}$ is a weight function on $G^{\boxtimes n}$. Similarly, there is an obvious weight function $p_1 + p_2$ defined on the disjoint union $G_1 + G_2$. When p_1 and p_2 are defined on the same graph, we use the same notation $p_1 + p_2$ for the pointwise sum; despite this ambiguous notation, the meaning will always be clear from the context.

DEFINITION A.3.1. *Let G be a graph equipped with vertex weights p.*

(a) The ***weighted independence number*** *$\alpha(G, p)$ is the largest total weight of an independent set in G.*

(b) The ***weighted Lovász number*** *$\vartheta(G, p)$ is*

$$\vartheta(G, p) \stackrel{def}{=} \min_{|\Psi\rangle,\, |\psi_v\rangle} \max_{v \in V} \frac{p(v)}{|\langle \Psi | \psi_v \rangle|^2} \tag{A.5}$$

where $|\Psi\rangle \in \mathbb{R}^{|V(G)|}$ ranges over all unit vectors and $(|\psi_v\rangle)_{v \in V(G)}$ over all orthonormal labelings.

(c) The ***weighted Shannon capacity*** *$\Theta(G, p)$ is*

$$\Theta(G, p) = \lim_{n \to \infty} \sqrt[n]{\alpha(G^{\boxtimes n}, p^{\otimes n})}. \tag{A.6}$$

(d) The ***weighted fractional packing number*** *$\alpha^*(G, p)$ is*

$$\alpha^*(G, p) \stackrel{def}{=} \max_q \sum_{v \in V} p(v) q(v).$$

where $q : V(G) \to \mathbb{R}_+$ ranges over all vertex weights satisfying $\sum_{v \in C} q(v) \leqslant 1$ for all cliques $C \subseteq V(G)$.

The fraction in (A.5) uses the convention $\frac{0}{0} = 0$. See [**Knu94**] for several equivalent definitions of $\vartheta(G, p)$.

All of these quantities specialize to their unweighted counterparts by choosing unit weights $p = \mathbb{1}$.

PROPOSITION A.3.2. *Let $\mathrm{Cl}(G)$ denote the set of all cliques on G.*

$$\alpha^*(G, p) = \min_x \sum_{C \in \mathrm{Cl}(G)} x(C) \tag{A.7}$$

where x ranges over all functions $x : \mathrm{Cl}(G) \to \mathbb{R}_+$ with $p(v) \leqslant \sum_{C \ni v} x(C)\ \forall v$.

PROOF. Linear programming duality. □

LEMMA A.3.3. *(a)*

$$\Theta(G_1 + G_2, p_1 + p_2) \geqslant \Theta(G_1, p_1) + \Theta(G_2, p_2). \tag{A.8}$$

(b)

$$\Theta(G_1 \boxtimes G_2, p_1 \otimes p_2) \geqslant \Theta(G_1, p_1)\Theta(G_2, p_2). \tag{A.9}$$

PROOF. As in the unweighted case [**Sha56**]. □

Since these inequalities are not tight in general in the unweighted case [**Hae79**, **Alo98**], neither can they be tight in the weighted case. One might expect simpler counterexamples to exist in the weighted case, but we have not been successful in finding any.

When p_1, p_2 are weight functions on the same graph G, superadditivity no longer holds for trivial reasons: e.g. for $G = K_2$, the graph on two adjacent vertices $\{u, v\}$ with $p_1 = \mathbb{1}_u$ and $p_2 = \mathbb{1}_v$, we have

$$1 = \Theta(G, p_1 + p_2) < \Theta(G, p_1) + \Theta(G, p_2) = 2.$$

Many statements about these invariants can be reduced to statements about their unweighted counterparts using a technique we call **blow-up**. Applying this technique requires the vertex weights to be rational. Therefore, we begin by proving a continuity result which allows us to reduce many problems to the case of rational weights.

LEMMA A.3.4. *Let (G, p) be a weighted graph and $\overline{K}_m$ the empty graph on m vertices with weights q. Then,*

$$X(G + \overline{K}_m, p + q) = X(G, p) + \sum_{v \in V(\overline{K}_m)} q(v) \tag{A.10}$$

for all four invariants $X \in \{\alpha, \Theta, \vartheta, \alpha^*\}$.

PROOF. This is trivial for $X = \alpha$. For $X = \vartheta$, it is a special case of [**Knu94**, eq. (18.2)]. For $X = \alpha^*$, it follows from an application of Proposition A.3.2. It remains to treat the case $X = \Theta$.

Since $\Theta(\overline{K}_m, q) = \sum_v q_v$, the inequality "$\geqslant$" is an instance of superadditivity (A.8) of Θ. To also show "$\leqslant$", we choose any independent set I in $(G + \overline{K}_m)^{\boxtimes n}$ and partition it into a disjoint union

$$I = \bigcup_{\vec{s} \in \{0,1\}^n} I_{\vec{s}}$$

where each $I_{\vec{s}}$ contains only vertices $(v_1, \ldots, v_n)$ with $v_i \in V(G)$ if $s_i = 0$ and $v_i \in V(\overline{K}_m)$ if $s_i = 1$. Then upon dropping all components i with $s_i = 1$, such an $I_{\vec{s}}$ becomes an independent set in some $G^{\boxtimes k}$. In this way, we get the estimate

$$\begin{aligned} \alpha\left((G + \overline{K}_m)^{\boxtimes n}, (p+q)^{\otimes n}\right) &\leqslant \sum_{k=0}^{n} \binom{n}{k} \alpha(G^{\boxtimes k}, p^{\otimes k}) \left(\sum_i q_i\right)^{n-k} \\ &\leqslant \sum_{k=0}^{n} \binom{n}{k} \Theta(G, p)^k \left(\sum_i q_i\right)^{n-k} = \left(\Theta(G, p) + \sum_i q_i\right)^n, \end{aligned}$$

which implies the desired inequality upon taking the n-th root and then $n \to \infty$. □

LEMMA A.3.5. *Let (G, p) be a weighted graph, $v \in G$ a vertex, $q \in \mathbb{R}_+$ and $X \in \{\alpha, \Theta, \vartheta, \alpha^*\}$. Then*

$$X(G, p) \leqslant X(G, p + q\mathbb{1}_v) \leqslant X(G, p) + q. \tag{A.11}$$

PROOF. The first inequality is clear since $X(G, p)$ is a non-decreasing function of p.

Since adding additional edges cannot increase the value of X and two vertices with exactly the same neighbors can be identified to one vertex by adding the weights (for $X = \vartheta$, see [**Knu94**, Lemma 16]), we have $X(G, p + q\mathbb{1}_v) \leqslant X(G + \overline{K}_1, p + q)$. Now the second inequality follows from the previous lemma with $m = 1$. □

This lemma directly gives the desired continuity result:

COROLLARY A.3.6. *For any graph G and any $X \in \{\alpha, \Theta, \vartheta, \alpha^*\}$, the function $p \mapsto X(G, p)$ is continuous.*

We can now introduce the blow-up technique which can be used to translate problems from the weighted case to the unweighted setting.

DEFINITION A.3.7. *Let (G, p) be a weighted graph with $p(v) \in \mathbb{N}\ \forall v$. Then the **blow-up** $\mathrm{Blup}(G, p)$ is the unweighted graph with vertex set*

$$\big\{(v, k)\ :\ v \in G, k \in \{1, \ldots, p(v)\}\big\},$$

where we take (v, k) and (v', k') to be adjacent if and only if $v \sim v'$ in G.

Intuitively speaking, $\mathrm{Blup}(G, p)$ is constructed by replacing every vertex v in G by $p(v)$ many non-adjacent vertices. In particular, if $p(v) = 0$, the vertex v simply gets removed from the graph. Blow-ups have also been considered in [**Knu94**, Sec. 16], although not under that name.

LEMMA A.3.8. *For vertex weights in $\mathbb{N}$,*

(a) $\mathrm{Blup}(G_1 + G_2, p_1 + p_2) = \mathrm{Blup}(G_1, p_1) + \mathrm{Blup}(G_2, p_2)$.
(b) $\mathrm{Blup}(G_1 \boxtimes G_2, p_1 \otimes p_2) = \mathrm{Blup}(G_1, p_1) \boxtimes \mathrm{Blup}(G_2, p_2)$*;*
(c) $X(\mathrm{Blup}(G, p)) = X(G, p)$ *for every* $X \in \{\alpha, \Theta, \vartheta, \alpha^*\}$.

PROOF. Clear. □

We can now already reap some of the simpler benefits of the blow-up technique:

COROLLARY A.3.9.

$$\alpha(G, p) \leqslant \Theta(G, p) \leqslant \vartheta(G, p) \leqslant \alpha^*(G, p).$$

PROOF. Combine Lemma A.3.8 with Proposition A.1.11. □

COROLLARY A.3.10 ([**Knu94**, (20.5)]).

$$\vartheta(G_1 \boxtimes G_2, p_1 \otimes p_2) = \vartheta(G_1, p_1)\vartheta(G_2, p_2)$$

PROOF. Combine Lemma A.3.8 with Proposition A.1.9. □

COROLLARY A.3.11.

$$\alpha(G_1 \boxtimes G_2, p_1 \otimes p_2) \geqslant \alpha(G_1, p_1)\alpha(G_2, p_2)$$

PROOF. Combine Lemma A.3.8 with Lemma A.1.2. □

A.4. Main conjectures for weighted graphs. We now extend the material of Section A.2 to the weighted case.

DEFINITION A.4.1. *A weighted graph (G, p) is **single-shot** if $\alpha(G, p) = \Theta(G, p)$.*

CONJECTURE A.4.2. *For all weighted single-shot graphs (G_1, p_1), (G_2, p_2),*

(a) $\alpha(G_1 \boxtimes G_2, p_1 \otimes p_2) = \alpha(G_1, p_2)\alpha(G_2, p_2)$*;*
(b) $\Theta(G_1 \boxtimes G_2, p_1 \otimes p_2) = \Theta(G_1, p_1)\Theta(G_2, p_2)$*;*
(c) $(G_1 + G_2, p_1 + p_2)$ *is also single-shot, so that* $\Theta(G_1 + G_2, p_1 + p_2) = \Theta(G_1, p_1) + \Theta(G_2, p_2)$.

Applying the blow-up technique in order to relate these conjectures to their unweighted counterparts requires some further preparation.

LEMMA A.4.3. *Let (G, p) be a weighted single-shot graph. Then for every $\varepsilon > 0$ there exist weights $p'(v) \in \mathbb{Q}$ with $|p(v) - p'(v)| < \varepsilon$ and such that (G, p') is still single-shot with $\alpha(G, p') = \alpha(G, p)$.*

Proof. Let $p_{\max}$ be the largest weight of a vertex in G, and fix $\delta > 0$ such that $2\delta \cdot p_{\max} \leqslant \varepsilon$. Fix any independent set $v_1, \ldots, v_n$ of maximal weight and choose rational numbers $p'(v_i) \in \big((1-\delta)p(v_i), (1+\delta)p(v_i)\big)$ such that $\sum_i p'(v_i) = \sum_i p_i = \alpha(G,p)$. Furthermore, for vertices w not in that set, choose arbitrary rational numbers $p'(w) \in \big((1-2\delta)p(w), (1-\delta)p(w)\big)$. Then $2\delta \cdot p_{\max} \leqslant \varepsilon$ guarantees $|p(v) - p'(v)| < \varepsilon$ for all $v \in V(G)$.

Now we claim that $\alpha(G,p') = \Theta(G,p') = \alpha(G,p)$. Upon setting $q_i \stackrel{def}{=} p'(v_i) - (1-\delta)p(v_i)$, we estimate

$$\alpha(G,p') \leqslant \Theta(G,p') \leqslant \Theta\left(G,(1-\delta)p\right) + \sum_i q_i,$$

where the last inequality follows from Lemma A.3.4 and the fact that transporting some weight from some vertex to a new isolated vertex cannot decrease the capacity. Since $\sum_i q_i = \alpha(G,p) - (1-\delta)\alpha(G,p)$, we can further evaluate this to

$$\alpha(G,p') \leqslant \Theta(G,p') \leqslant (1-\delta)\Theta(G,p) + \delta\alpha(G,p) = \alpha(G,p).$$

On the other hand, we have constructed p' in such a way that there is an independent set of weight $\alpha(G,p)$, and hence all these inequalities are actually equalities. □

Theorem A.4.4. *Each one of the Conjectures A.4.2 is equivalent to its sibling in A.2.1. In particular, Conjectures A.4.2(a) and A.4.2(b) are equivalent and imply A.4.2(c).*

Proof. By taking all weights to be 1, each statement in Conjecture A.2.1 becomes a special case of the corresponding statement in Conjecture A.4.2. Therefore, the bulk of the proof lies in showing the converse implications.

We illustrate how to translate any potential counterexample $(G_1 \boxtimes G_2, p_1 \otimes p_2)$ to Conjecture A.4.2(a) into a counterexample to Conjecture A.2.1(a). To this end, we first apply Lemma A.4.3 to both (G_j, p_j) with a certain $\varepsilon > 0$ and obtain (G_j, p'_j). Then, the difference $(p'_1 \otimes p'_2)(v_1, v_2) - (p_1 \otimes p_2)(v_1, v_2)$ can be bounded by a certain function of ε and the $\alpha(G_j, p_j)$'s. In particular, one can choose ε so small that

$$\alpha(G_1 \boxtimes G_2, p'_1 \otimes p'_2) > \alpha(G_1, p_1)\alpha(G_2, p_2) = \alpha(G_1, p'_1)\alpha(G_1, p'_2).$$

After multiplying both weight functions p'_j by the respective common denominator, they become integer-valued, and the claim then follows from Lemma A.3.8.

Conjecture A.2.1(b) follows from A.4.2(b) by analogous reasoning, now also relying on Proposition A.3.6 for $X = \Theta$. The same holds for the implication of Conjecture A.2.1(c) from A.4.2(c). □

Conjecture A.4.5. *Let G be a graph with weight functions p_1 and p_2 such that both (G, p_1) and (G, p_2) are single-shot. Then*

$$\Theta(G, p_1 + p_2) \leqslant \Theta(G, p_1) + \Theta(G, p_2).$$

Proposition A.4.6. *Conjecture A.4.5 is equivalent to A.2.1(c).*

Proof. We prove equivalence to Conjecture A.4.2(c), and then the claim follows from Theorem A.4.4.

First, Conjecture A.4.2(c) easily follows from A.4.5 by taking $G = G_1 + G_2$ and using (A.8). Hence it remains to prove that A.4.2(c) implies A.4.5. To see this, consider the graph G' with vertex set

$$\left\{ (v,k) \ : v \in G, k \in \{1,2\} \right\}$$

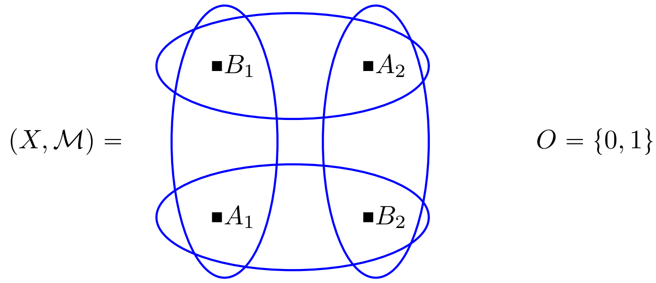


FIGURE 11. The CHSH scenario as a marginal scenario. We now draw the vertices as squares in order to indicate that the interpretation differs from the one of all other illustrations of hypergraphs in this paper.

where we take (v, k) and (v', k') to be adjacent if and only if $v \sim v'$ in G. We equip G' with the weight function p' given by $p'(v, 1) \stackrel{def}{=} p_1(v)$ and $p'(v, 2) \stackrel{def}{=} p_2(v)$. Then, by construction, $\Theta(G', p') = \Theta(G, p_1 + p_2)$.

Now this (G', p') is a union of (G, p_1) and (G, p_2), which it contains as induced weighted subgraphs. Since removing edges cannot decrease Θ, we get

$$\Theta(G, p_1 + p_2) = \Theta(G', p') \leqslant \Theta(G + G, p_1 + p_2) = \Theta(G, p_1) + \Theta(G, p_2).$$

as was to be shown. □

Appendix B. Relation to the observable-based approach

The observable-based approach to quantum contextuality and nonlocality has first been studied explicitly by Abramsky and Brandenburger [**AB11**]. It was used much earlier in a different mathematical context by Vorob'ev [**Vor62**]. See also [**LSW11**, **FC12**], where similar definitions have been used. In this section, our goal is to show how the observable-based approach can be embedded into our formalism. A converse construction should be possible upon augmenting the observable-based approach by additional constraints as in [**AB11**, Sec. 7]. In this sense, the two formalism are essentially equivalent. We believe that both approaches have their merits; for example, in both cases the relation to sophisticated mathematical methods can be exploited. In the observable-based approach, this has been done in [**AMSB11**]; for the hypergraph-based approach, this has been started in [**CSW10**] and further developed in this paper.

B.1. Definitions for the observable-based approach. The following definition blends the terminology of [**AB11**] with the one of [**FC12**].

DEFINITION B.1.1. *A* ***marginal scenario*** $(X, O, \mathcal{M})$ *is a finite set* $X = \{A_1, \ldots, A_n\}$*, the elements of which we call* ***observables****, together with a finite set* O *of outcomes and a* ***measurement cover*** $\mathcal{M}$*, which is a family of subsets* $\mathcal{M} \subseteq 2^X$ *such that*

(a) every element of X *occurs in some* C*:* $\bigcup_{C \in \mathcal{M}} C = X$*.*
(b) $\mathcal{M}$ *is an anti-chain:* $C, C' \in \mathcal{M}$*,* $C \subseteq C' \implies C = C'$*.*

The $C \in \mathcal{M}$ *are called* ***measurement contexts****.*

From the mathematical point of view, the maximal sets of compatible observables are a hypergraph precisely as in Definition 2.2.1, but the physical interpretation is quite different. The subsets

in $\mathcal{M}$ represent the maximal sets of jointly measurable observables. See Figure 11 for an example in which the four pairs

$$\{A_1, B_1\}, \quad \{A_1, B_2\}, \quad \{A_2, B_1\}, \quad \{A_2, B_2\}$$

are jointly measurable, but no other pairs or triples of observables are jointly measurable.

As is common practice with many other mathematical structures, we denote a marginal scenario $(X, O, \mathcal{M})$ simply by X, at least when O and $\mathcal{M}$ are clear from the context.

As noted in [**AB11**], it is not a substantial restriction to assume that all observables take values in the same set of outcomes O. We assume this mainly for convenience of notation and note that all of our considerations and results can easily be extended to the general case in which each measurement $A \in X$ takes values in an associated finite set of outcomes O_A depending on A.

In the following, we want to consider measurements of compatible observables which are conducted in a certain temporal order. Assume that we have already measured some observable $A \in X$; then is it possible to define a marginal scenario which encodes all the possibilities for subsequent measurements? The following notion achieves this:

Definition B.1.2. *Given an observable $A \in X$, the **induced marginal scenario** $X\{A\}$ is the marginal scenario having observables*

$$X\{A\} = \left\{ A' \in X \,\middle|\, A' \neq A,\, \exists C \in \mathcal{M} \text{ s.t. } \{A, A'\} \subseteq C \right\}$$

and measurement contexts defined to be the restrictions of those $C \in \mathcal{M}$ with $A \in C$ down to $X\{A\}$.

By definition, any $X\{A\}$ has a smaller number of observables than the original X. In particular, iterating this construction by taking an induced marginal scenario of an induced marginal scenario etc., one eventually ends up with an empty scenario, and the process terminates.

We make the following recursive definition:

Definition B.1.3. *A **measurement protocol** T on a marginal scenario X is*

(a) $T = \varnothing$ if $X = \varnothing$;
(b) otherwise, $T = (A, f)$, where $A \in X$ is an observable and $f : O \to \mathrm{MP}(X\{A\})$ is a function, where $\mathrm{MP}(X\{A\})$ is the set of all measurement protocols on the scenario $X\{A\}$.

Intuitively, a measurement protocol consists of a choice of observable and an assignment of a new measurement protocol to each outcome of the observable, where the new measurement protocol lives on the induced marginal scenario.

Upon unraveling the recursive structure of this definition, one finds that a measurement protocol specifies sequences of measurements which can be applied to the system, where the choices of subsequent measurements f are allowed to depend on the outcomes of the earlier ones. These measurement sequences have the additional property that all measurements in a sequence are compatible and that no measurement can occur twice in the same sequence. We use the letter "T" to indicate the tree-like appearance of this structure. Note that every measurement sequence is automatically maximal in the sense that it contains all observables of a certain measurement context.

The set of outcomes $\mathrm{Out}(T)$ of a measurement protocol T is also defined recursively: if $T = \varnothing$, then there is only a single outcome which we denote by "$*$", so that $\mathrm{Out}(\varnothing) = \{*\}$. Otherwise, we have $T = (A, f)$ and put

$$\mathrm{Out}(T) \stackrel{def}{=} \{\, (a, \alpha) \,:\, a \in O,\ \alpha \in \mathrm{Out}(f(a)) \,\}\,.$$

In this way, an element of $\mathrm{Out}(T)$ corresponds to a measurement sequence in T together with an associated sequence of outcomes for these measurements such that applying the protocol to any

outcome in the sequence results in the following measurement, except if the outcome is the last one in the sequence.

Definition B.1.4. *The contextuality scenario $H[X]$ associated to a marginal scenario X has vertices*

$$V(H[X]) \stackrel{def}{=} \left\{s \in O^C \;:\; C \in \mathcal{M}\right\}$$

and edges

$$E(H[X]) \stackrel{def}{=} \{\mathrm{Out}(T) \;:\; T \in \mathrm{MP}(X)\}.$$

We write P for an **empirical model** on X [**AB11**]. This means that for each $C \in \mathcal{M}$, P_C is a probability distribution over O^C, such that the **sheaf condition** holds:

$$P_{C|C\cap C'} = P_{C'|C\cap C'} \quad \forall C, C' \in \mathcal{M}, \tag{B.1}$$

where $P_{C|C\cap C'}$ stands for the marginal distribution of P_C associated to the observables in $C \cap C'$. For an assigment of outcomes $s \in O^C$, $P_C(s)$ is to be thought of as the probability of obtaining the joint outcome s when jointly measuring all observables in C. The sheaf condition is a generalization of the no-signaling condition.

B.2. Correspondence to our approach. To an empirical model P we associate a probabilistic model on $H[X]$ by setting, for each $C \in \mathcal{M}$ and each $s \in O^C$,

$$p(s : C \to O) \stackrel{def}{=} P_C(s). \tag{B.2}$$

It will need to be verified that this actually is a probabilistic model, i.e. that these probabilities are suitably normalized for every edge in $E[X]$.

Conversely, given a probabilistic model p on $H[X]$, we claim that (B.2) defines an empirical model P on X.

Theorem B.2.1. *This defines a linear bijection between empirical models on X and probabilistic models on $H[X]$.*

This bijective correspondence generalizes Proposition 3.3.2 and the related [**FSA⁺12**, Lemma 1].

Proof. We first verify that (B.2) turns an empirical model P into a probabilistic model p. It needs to be shown that

$$\sum_{s\in\mathrm{Out}(T)} P_C(s) = 1 \tag{B.3}$$

for any measurement protocol T. In order to prove this, we introduce the notion of **post-measurement** empirical model. Suppose that a measurement has resulted in an outcome $a \in O$ for an observable $A \in X$. Then for the subsequent measurements in the scenario $X\{A\}$, we expect the posterior probabilities

$$P_C^{\mathrm{post}(a)}(s) = \frac{P_C(s)}{P_{\{A\}}(a)}.$$

We now use induction on the size of X in order to prove (B.3). The base case is $X = \varnothing$, in which there is nothing to prove. For the induction step, we decompose $T = (A, f)$ and use the induction assumption on each $P^{\mathrm{post}(a)}$ for those $a \in O$ with $P_{\{A\}}(a) \neq 0$. Then

$$\sum_{s\in\mathrm{Out}(T)} P_C(s) = \sum_a \sum_{\alpha\in\mathrm{Out}(f(a))} P_{\{A\}}(a) P_C^{\mathrm{post}(a)}(\alpha) = \sum_a P_{\{A\}}(a) = 1,$$

as was to be shown.

Conversely, we need to prove that if p is a probabilistic model on $H[X]$, then the associated P is an empirical model, i.e. that it satisfies (B.1). It is sufficient to consider the case $C\cap C'\neq\varnothing$, for otherwise (B.1) is vacuous. Let $s_0\in O^{C\cap C'}$ be an arbitrary joint outcome of the observables $C\cap C'$. Then we consider a measurement protocol T given by conducting the measurements $C\cap C'$, and then conducting the measurements $C\backslash C'$ if the joint outcome was s_0, and conducting the measurements $C'\backslash C$ otherwise. Then the normalization equation associated to this measurement protocol reads

$$\sum_{t\in O^{C\backslash C'}} p(s_0\cup t)+\sum_{s_0\neq s\in O^{C\cap C'}}\sum_{t'\in O^{C'\backslash C}} p(s\cup t')=1.$$

Comparing this with the normalization equation associated to the measurement protocol which simply measures all observables in C' and outputs their joint outcome,

$$\sum_{s\in O^{C\cap C'}}\sum_{t'\in O^{C'\backslash C}} p(s'\cup t')=1,$$

gives, upon splitting the latter equation into the $s=s_0$ part and the $s\neq s_0$ part,

$$\sum_{t\in O^{C\backslash C'}} p(s_0\cup t)=\sum_{t'\in O^{C'\backslash C}} p(s_0\cup t'),$$

as was to be shown. □

There are analogous correspondence theorems for quantum models and classical models. Since these are perfectly analogous both in the statement and in the proof, we do not discuss them further.

References

[AB11] Samson Abramsky and Adam Brandenburger, *The sheaf-theoretic structure of non-locality and contextuality*, New Journal of Physics **13** (2011), no. 11, 113036. ↑3, 5, 7, 17, 53, 54, 55

[AII06] David Avis, Hiroshi Imai, and Tsuyoshi Ito, *On the relationship between convex bodies related to correlation experiments with dichotomic observables*, J. Phys. A **39** (2006), 11283. ↑38

[AL06] Noga Alon and Eyal Lubetzky, *The Shannon capacity of a graph and the independence numbers of its powers*, IEEE Trans. Inf. Theory **52** (2006), no. 5. ↑46

[Alo98] Noga Alon, *The Shannon capacity of a union*, Combinatorica **18** (1998), no. 3, 301–310. ↑46, 48, 49

[AMSB11] Samson Abramsky, Shane Mansfield, and Rui Soares Barbosa, *The Cohomology of Non-Locality and Contextuality*, Proceedings 8th International Workshop on Quantum Physics and Logic (Nijmegen, 2011), 2011. ↑53

[Anj04] Miguel F. Anjos, *On semidefinite programming relaxations for the satisfiability problem*, Math. Methods Oper. Res. **60** (2004), no. 3, 349–367. ↑37

[AQB+12] Mateus Araújo, Marco Túlio Quintino, Costantino Budroni, Marcelo Terra Cunha, and Adán Cabello, *Complete characterization of the n-cycle noncontextual polytope*, 2012. arXiv:1206.3212. ↑41

[Bar07] Jonathan Barrett, *Information processing in generalized probabilistic theories*, Phys. Rev. A **75** (2007), no. 3, 032304. ↑7

[Bel64] John S. Bell, *On the Einstein-Podolsky-Rosen paradox*, Physics **1** (1964), 195–200. ↑3, 17

[Ber61] C. Berge, *Färbung von Graphen, deren sämtliche bzw. deren ungerade Kreise starr sind*, Wiss. Z. Martin-Luther-Univ. Halle-Wittenberg Math.-Natur. Reihe **10** (1961), no. 114. ↑34

[Ber97] Claude Berge, *Motivations and history of some of my conjectures*, Discrete Math. **165–166** (1997), 61–70. ↑46

[BFRW05] Howard Barnum, Christopher A. Fuchs, Joseph M. Renes, and Alexander Wilce, *Influence-free states on compound quantum systems*, 2005. arXiv:quant-ph/0507108. ↑11, 12

[BL09] Jonathan Barrett and Matthew Leifer, *The de Finetti theorem for test spaces*, New Journal of Physics **11** (2009), no. 3, 033024. ↑11

[BPT11] Christine Bachoc, Arnaud Pécher, and Alain Thiéry, *On the theta number of powers of cycle graphs*, 2011. arXiv:1103.0444. ↑43

[Cab08] Adán Cabello, *Experimentally Testable State-Independent Quantum Contextuality*, Phys. Rev. Lett. **101** (2008), no. 21, 210401. ↑5, 6

[Cab12a] ———, *Twin inequality for fully contextual quantum correlations*, 2012. arXiv:1209.0112. ↑43

[Cab12b] ———, *A simple explanation of the quantum violation of a fundamental inequality*, 2012. arXiv:1210.2988. ↑31, 43

[Cab12c] ———, *Specker's fundamental principle of quantum mechanics*, 2012. arXiv:1212.1756. ↑4, 29, 30

[CDLTP12] Adán Cabello, Lars Eirik Danielsen, Antonio J. López-Tarrida, and José R. Portillo, *Basic logical structures in quantum correlations*, 2012. arXiv:1211.5825. ↑35

[CEGA96] Adán Cabello, José M. Estebaranz, and Guillermo García-Alcaine, *Bell-Kochen-Specker theorem: a proof with* 18 *vectors*, Phys. Lett. A **212** (1996), no. 4, 183–187. ↑5, 6, 17, 22

[CF12] Rafael Chaves and Tobias Fritz, *Entropic approach to local realism and noncontextuality*, Phys. Rev. A **85** (2012), no. 3, 032113. ↑7

[CGR03] B. Codenotti, I. Gerace, and G. Resta, *Some remarks on the Shannon capacity of odd cycles*, Ars Combinatoria **66** (2003), 243–257. ↑46

[CHSH69] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt, *Proposed Experiment to Test Local Hidden-Variable Theories*, Phys. Rev. Lett. **23** (1969), no. 15, 880–884. ↑5

[CMW00] Bob Coecke, David Moore, and Alexander Wilce, *Operational quantum logic: an overview*, Current research in operational quantum logic, 2000, pp. 1–36. ↑3

[CRST06] Maria Chudnovsky, Neil Robertson, Paul Seymour, and Robin Thomas, *The strong perfect graph theorem*, Ann. of Math. (2) **164** (2006), no. 1, 51–229. ↑35

[CSW10] Adán Cabello, Simone Severini, and Andreas Winter, *(Non-)Contextuality of Physical Theories as an Axiom*, 2010. arXiv:1010.2163. ↑3, 5, 21, 27, 30, 35, 39, 43, 53

[D'A10] Giacomo Mauro D'Ariano, *Probabilistic theories: what is special about quantum mechanics?*, Philosophy of Quantum Information and Entanglement, 2010. ↑12

[EF70] J. Edmonds and D.R. Fulkerson, *Bottleneck extrema*, Journal of Combinatorial Theory **8** (1970), no. 3, 299–306. ↑6

[Eit94] Thomas Eiter, *Exact transversal hypergraphs and application to Boolean μ-functions*, J. Symbolic Comput. **17** (1994), no. 3, 215–225. ↑18, 38

[Eng97] Konrad Engel, *Sperner theory*, Encyclopedia of Mathematics and its Applications, vol. 65, Cambridge University Press, Cambridge, 1997. ↑6

[FC12] Tobias Fritz and Rafael Chaves, *Entropic Inequalities and the Marginal Problem*, IEEE Trans. Inf. Theory (2012). arXiv:1112.4788. To appear. ↑7, 53

[Fek23] Michael Fekete, *Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten*, Math. Z. **17** (1923), no. 1, 228–249. ↑46

[Fin82] Arthur Fine, *Hidden variables, joint probability, and the Bell inequalities*, Phys. Rev. Lett. **48** (1982), no. 5, 291–295. ↑17

[FGR92] D. J. Foulis, R. J. Greechie, and G. T. Rüttimann, *Filters and supports in orthoalgebras*, Internat. J. Theoret. Phys. **31** (1992), no. 5, 789–807. ↑40

[FNT12] Tobias Fritz, Tim Netzer, and Andreas Thom, *Can you compute the operator norm?*, 2012. arXiv:1207.0975. ↑37, 38

[FR72] D. J. Foulis and C. H. Randall, *Operational statistics. I. Basic concepts*, J. Mathematical Phys. **13** (1972), 1667–1675. ↑7, 18

[FR81] ———, *Empirical logic and tensor products*, Interpretations and foundations of quantum theory (Marburg, 1979), 1981, pp. 9–20. ↑10, 11

[Fri12] Tobias Fritz, *Tsirelson's problem and Kirchberg's conjecture*, Rev. Math. Phys. **24** (2012), 1250012. ↑21, 22, 38

[FSA+12] Tobias Fritz, Ana Belén Sainz, Remigiusz Augusiak, Jonatan Bohr Brask, Rafael Chaves, Anthony Leverrier, and Antonio Acín, *Local orthogonality: a multipartite principle for correlations*, 2012. arXiv:1210.3018. ↑4, 5, 9, 30, 31, 32, 43, 55

[GHSZ90] Daniel M. Greenberger, Michael A. Horne, Abner Shimony, and Anton Zeilinger, *Bell's theorem without inequalities*, Amer. J. Phys. **58** (1990), no. 12, 1131–1143. ↑22

[GLS01] G. Gottlob, N. Leone, and F. Scarcello, *Hypertree decompositions: A survey*, Mathematical Foundations of Computer Science 2001 (2001), 37–57. ↑18

[Gre71] R. J. Greechie, *Orthomodular lattices admitting no states*, J. Combinatorial Theory Ser. A **10** (1971), 119–132. ↑7

[GP79] Martin Grötschel and Manfred W. Padberg, *On the symmetric travelling salesman problem. I. Inequalities*, Math. Programming **16** (1979), no. 3, 265–280. ↑44

[Haa96] Rudolf Haag, *Local quantum physics*, Second, Texts and Monographs in Physics, Springer-Verlag, Berlin, 1996. ↑3

[Hae79] Willem Haemers, *On some problems of Lovász concerning the Shannon capacity of a graph*, IEEE Trans. Inform. Theory **25** (1979), no. 2, 231–232. ↑46, 47, 48, 49

[Hae81] ———, *An upper bound for the Shannon capacity of a graph*, Algebraic methods in graph theory, Vol. I, II (Szeged, 1978), 1981, pp. 267–272. ↑32, 33, 48

[Hen12] Joe Henson, *Quantum contextuality from a simple principle?*, 2012. arXiv:1210.5978. ↑4, 5, 29

[IK00] Wilfried Imrich and Sandi Klavžar, *Product graphs: structure and recognition*, Wiley-Intersciene, New York, 2000. ↑45

[JNP+11] Marius Junge, Miguel Navascués, Carlos Palazuelos, David Pérez-García, Volkher B. Scholz, and Reinhard F. Werner, *Connes embedding problem and Tsirelson's problem*, J. Math. Phys. **52** (2011), no. 1, 012102, 12. ↑21, 22

[Kar72] Richard M. Karp, *Reducibility among combinatorial problems*, Complexity of computer computations (Proc. Sympos., IBM Thomas J. Watson Res. Center, Yorktown Heights, N.Y., 1972), 1972, pp. 85–103. ↑36

[KCBS08] A. A. Klyachko, M. A. Can, S. Binicioglu, and A. S. Shumovsky, *A simple test for hidden variables in spin-1 system*, Phys. Rev. Lett. **101** (2008), 020403–020406. ↑35, 40, 41

[Knu94] Donald E. Knuth, *The Sandwich Theorem*, Electron. J. Comb. **1** (1994), A1. ↑34, 38, 43, 47, 49, 50, 51

[KR83] Richard V. Kadison and John R. Ringrose, *Fundamentals of the theory of operator algebras. Vol. I*, Pure and Applied Mathematics, vol. 100, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1983. Elementary theory. ↑25

[KS67] Simon Kochen and E. P. Specker, *J. Math. Mech.* **17** (1967), 59–87. ↑3, 17

[Las02] Jean B. Lasserre, *An explicit equivalent positive semidefinite program for nonlinear* 0-1 *programs*, SIAM J. Optim. **12** (2002), no. 3, 756–769 (electronic). ↑27, 36

[Lau03] Monique Laurent, *A comparison of the Sherali-Adams, Lovász-Schrijver, and Lasserre relaxations for 0-1 programming*, Math. Oper. Res. **28** (2003), no. 3, 470–496. ↑37

[Lov72] László Lovász, *Normal hypergraphs and the perfect graph conjecture*, Discrete Math. **2** (1972), no. 3, 253–267. ↑34

[Lov78] ———, *Kneser's conjecture, chromatic number, and homotopy*, J. Combin. Theory Ser. A **25** (1978), no. 3, 319–324. ↑44

[Lov79] ———, *On the Shannon capacity of a graph*, IEEE Trans. Inform. Theory **25** (1979), no. 1, 1–7. ↑47

[LSW11] Yeong-Cherng Liang, Robert W. Spekkens, and Howard M. Wiseman, *Specker's parable of the overprotective seer: a road to contextuality, nonlocality and complementarity*, Phys. Rep. **506** (2011), no. 1-2, 1–39. ↑3, 7, 17, 30, 53

[NP01] Denis Naddef and Yves Pochet, *The symmetric traveling salesman polytope revisited*, Math. Oper. Res. **26** (2001), no. 4, 700–722. ↑44

[NPA07] Miguel Navascués, Stefano Pironio, and Antonio Acín, *Bounding the Set of Quantum Correlations*, Phys. Rev. Lett. **98** (2007), no. 1, 010401. ↑5, 23, 25

[NPA08] Miguel Navascués, Stefano Pironio, and Antonio Acín, *A convergent ierarchy of semidefinite programs characterizing the set of quantum correlations*, New Journal of Physics **10** (2008), no. 7, 073013. ↑5, 23, 25, 27, 32

[PMMF10] Mladen Pavičić, Brendan D. McKay, Norman D. Megill, and Krešimir Fresl, *Graph approach to quantum systems*, J. Math. Phys. **51** (2010), no. 10, 102103, 31. ↑7

[PMMM05] Mladen Pavičić, Jean-Pierre Merlet, Brendan McKay, and Norman D. Megill, *Kochen-Specker vectors*, J. Phys. A **38** (2005), no. 7, 1577–1592. ↑43

[PNA10] Stefano Pironio, Miguel Navascués, and Antonio Ac´n, *Convergent Relaxations of Polynomial Optimization Problems with Noncommuting Variables*, SIAM Journal on Optimization **20** (2010), no. 5, 2157–2180. ↑5, 23, 27

[PR94] Sandu Popescu and Daniel Rohrlich, *Quantum nonlocality as an axiom*, Foundations of Physics **24** (1994), no. 3, 379–385. ↑14

[RF73] C. H. Randall and D. J. Foulis, *Operational statistics. II. Manuals of operations and their logics*, J. Mathematical Phys. **14** (1973), 1472–1480. ↑7

[Sch78] Thomas J. Schaefer, *The complexity of satisfiability problems*, Proceedings of the tenth annual ACM symposium on Theory of computing, 1978, pp. 216–226. ↑36

[Sch03] Alexander Schrijver, *Combinatorial optimization. Polyhedra and efficiency.*, Algorithms and Combinatorics, vol. 24, Springer-Verlag, Berlin, 2003. ↑36, 41, 43, 44

[Sha56] Claude E. Shannon, *The zero error capacity of a noisy channel*, Institute of Radio Engineers, Transactions on Information Theory, **IT-2** (1956), no. September, 8–19. ↑46, 49

[Shu74] Frederic W. Shultz, *A characterization of state spaces of orthomodular lattices*, J. Combinatorial Theory Ser. A **17** (1974), 317–328. ↑9, 35

[Spe60] Ernst Specker, *The logic of non-simultaneously decidable propositions*, 1960. Translation from the German original by M.P. Seevinck (2011), arXiv:1103.4537. ↑29

[Spe05] Robert W. Spekkens, *Contextuality for preparations, transformations, and unsharp measurements*, Phys. Rev. A **71** (2005), no. 5, 052108. ↑3, 6

[ST96] K. Svozil and J. Tkadlec, *Greechie diagrams, nonexistence of measures in quantum logics, and Kochen-Specker-type constructions*, J. Math. Phys. **37** (1996), no. 11. ↑7

[Svo93] Karl Svozil, *Randomness & undecidability in physics*, World Scientific Publishing Co. Inc., River Edge, NJ, 1993. ↑38

[Tar51] Alfred Tarski, *A decision method for elementary algebra and geometry*, University of California Press, Berkeley and Los Angeles, Calif., 1951. 2nd ed. ↑37

[Tka00] Josef Tkadlec, *Diagrams of Kochen-Specker type constructions*, Internat. J. Theoret. Phys. **39** (2000), no. 3, 921–926. Quantum structures '98 (Liptovský Ján). ↑7

[Tsi93] Boris S. Tsirelson, *Some results and problems on quantum Bell-type inequalities*, Hadronic Journal Supplement **8** (1993), 329–345. ↑14

[Vol09] Vitaly I. Voloshin, *Introduction to graph and hypergraph theory*, Nova Science Publishers Inc., New York, 2009. ↑43

[Vor62] N. N. Vorob'ev, *Consistent Families of Measures and Their Extensions*, Theory of Probability and its Applications **7** (1962), no. 2, 147-163. ↑53

[Wil05] Alexander Wilce, *Topological Test Spaces*, International Journal of Theoretical Physics **44** (2005), no. 8, 1227–1238. ↑15

[Wil08] ______, *Formalism and Interpretation in Quantum Theory*, 2008. A slightly edited version of a paper to appear as part of a Festchrift for Jeff Bub. ↑7, 12, 40

[Wil09] ______, *Test spaces*, Handbook of quantum logic and quantum structures—quantum logic, 2009, pp. 443–549. ↑3, 7

[Wri78] Ron Wright, *The state of the pentagon: a nonclassical example*, Mathematical foundations of quantum theory (Proc. Conf., Loyola Univ., New Orleans, La., 1977), 1978, pp. 255–274. ↑40

[Zeh06] H. Dieter Zeh, *Quantum nonlocality vs. Einstein locality*, 2006. http://www.rzuser.uni-heidelberg.de/~as3/nonlocality.html. ↑3

[Zie95] Günter M. Ziegler, *Lectures on Polytopes*, Graduate Texts in Mathematics, vol. 152, Springer Verlag, New York, 1995. ↑9

$$
\begin{tikzcd}
A.2.1(a) \arrow[r, Leftrightarrow, "A.4.4"] \arrow[d, Leftrightarrow, "A.2.2"'] & A.4.2(a) \arrow[r, Rightarrow, "7.5.2"] \arrow[d, Leftrightarrow, "A.4.4"'] & 7.5.1(a) \arrow[d, Leftrightarrow, "7.5.2"'] \\
A.2.1(b) \arrow[r, Leftrightarrow, "A.4.4"] \arrow[d, Rightarrow, "A.2.2"'] & A.4.2(b) \arrow[r, Rightarrow, "7.5.2"] \arrow[d, Rightarrow, "A.4.4"'] & 7.5.1(b) \arrow[dd, Rightarrow, "7.5.2"'] \\
A.2.1(c) \arrow[r, Leftrightarrow, "A.4.4"] & A.4.2(c) \arrow[dr, Rightarrow, "7.5.2"] \arrow[dd, Leftrightarrow, "A.4.6"'] & \\
& & 7.5.1(c) \\
& A.4.5 \arrow[ur, Rightarrow] &
\end{tikzcd}
$$

Figure 12. Our conjectures and the known implications between them.

Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada
E-mail address: tfritz@perimeterinstitute.ca

INRIA Rocquencourt, Domaine de Voluceau, B.P. 105, 78153 Le Chesnay Cedex, France
E-mail address: anthony.leverrier@inria.fr

ICFO–Institut de Ciencies Fotoniques, E–08860 Castelldefels, Barcelona, Spain
E-mail address: belen.sainz@icfo.es